Cell

Defining the Teratoma as a Model for Multi-lineage Human Development

Graphical Abstract



Authors

Daniella McDonald, Yan Wu, Amir Dailamy, ..., Ann Tipps, Kun Zhang, Prashant Mali

Correspondence

kzhang@bioeng.ucsd.edu (K.Z.), pmali@ucsd.edu (P.M.)

In Brief

The teratoma is characterized as a model for multi-lineage human development with cell types represented across all three germ layers and utilized to enable assaying of the effects of genetic perturbations simultaneously across multiple cell types, and a molecular sculpting strategy is presented to enrich for specific tissues.

Highlights

- Identified 20 preliminary teratoma cell types via scRNA-seq, histology, and RNA FISH
- Benchmarked teratoma brain and gut cell types against human fetal scRNA-seq datasets
- Demonstrated teratomas enable CRISPR screens across multiple cell types simultaneously
- Engineered teratomas with miRNA gene circuits to enrich for a specific lineage



Cell



Resource

Defining the Teratoma as a Model for Multi-lineage Human Development

Daniella McDonald,^{1,2,6} Yan Wu,^{1,6} Amir Dailamy,¹ Justin Tat,³ Udit Parekh,⁵ Dongxin Zhao,¹ Michael Hu,¹ Ann Tipps,⁴ Kun Zhang,^{1,2,*} and Prashant Mali^{1,2,7,*}

¹Department of Bioengineering, University of California, San Diego, San Diego, CA 92093, USA

²Biomedical Sciences Graduate Program, University of California, San Diego, San Diego, CA 92093, USA

³Department of Biological Sciences, University of California, San Diego, San Diego, CA 92093, USA

⁴School of Medicine, University of California, San Diego, San Diego, CA 92103, USA

⁵Department of Electrical and Computer Engineering, University of California, San Diego, San Diego, CA 92093, USA

⁶These authors contributed equally

⁷Lead Contact

*Correspondence: kzhang@bioeng.ucsd.edu (K.Z.), pmali@ucsd.edu (P.M.) https://doi.org/10.1016/j.cell.2020.10.018

SUMMARY

We propose that the teratoma, a recognized standard for validating pluripotency in stem cells, could be a promising platform for studying human developmental processes. Performing single-cell RNA sequencing (RNA-seq) of 179,632 cells across 23 teratomas from 4 cell lines, we found that teratomas reproducibly contain approximately 20 cell types across all 3 germ layers, that inter-teratoma cell type heterogeneity is comparable with organoid systems, and teratoma gut and brain cell types correspond well to similar fetal cell types. Furthermore, cellular barcoding confirmed that injected stem cells robustly engraft and contribute to all lineages. Using pooled CRISPR-Cas9 knockout screens, we showed that teratomas can enable simultaneous assaying of the effects of genetic perturbations across all germ layers. Additionally, we demonstrated that teratomas can be sculpted molecularly via microRNA (miRNA)-regulated suicide gene expression to enrich for specific tissues. Taken together, teratomas are a promising platform for modeling multi-lineage development, pan-tissue functional genetic screening, and tissue engineering.

INTRODUCTION

Current understanding of early human development relies heavily on inference from animal models. Model systems such as frogs (Vastag et al., 2011), fish (Farrell et al., 2018), and mice (Cao et al., 2019; Pijuan-Sala et al., 2019) have demonstrated that many features of early embryogenesis are evolutionarily conserved across species (Lin et al., 2009; Peter and Davidson, 2011; Royo et al., 2011). However, several aspects of development are highly species-specific, especially in neural development (Raff, 1996; Richardson et al., 1997; Fougerousse et al., 2000; Hodge et al., 2019) Although there have been studies of human embryonic development (Miller et al., 2014; Zhu et al., 2018), such studies are limited by a scarcity of relevant biological material and key ethical constraints. There has thus been a push to establish models specific to human development.

Human pluripotent stem cells (hPSCs), such as embryonic stem cells (ESCs) or induced PSCs (iPSCs), have been used as developmental models by directing differentiation of ESCs or iPSCs into various cell types. These studies have shed light on processes such as lineage bifurcation (Yao et al., 2017) and heterogeneity (Wang et al., 2017) during human neuronal development as well as the presence of discrete cell states during early

ESC differentiation (Jang et al., 2017). Additionally, perturbation screens in these cell culture models have looked at the key regulators of differentiation (Parekh et al., 2018) and reprogramming (Tsunemoto et al., 2018). However, true human development takes place in 3 dimensions, which is difficult to capture with a 2-dimensional monolayer (Brown et al., 2018; Liu et al., 2018).

In this regard, newer methods for modeling human development use organoid systems. Organoids are 3D "mini-organs" derived from hPSCs in which the cells spontaneously selfassemble into differentiated, functional cell types that structurally and functionally mimic their in vivo counterparts (Huch and Koo, 2015; Clevers, 2016; Yin et al., 2016; Dutta et al., 2017; Huch et al., 2017; Fligor et al., 2018; Capowski et al., 2019; Collin et al., 2019). Use of organoids has enabled researchers to model human-specific development in a 3D context, which is especially beneficial for modeling rare genetic diseases or cancers (Dekkers et al., 2013; Bigorgne et al., 2014; Gao et al., 2014; Bartfeld et al., 2015; Boj et al., 2015; Huch and Koo, 2015; van de Wetering et al., 2015; Yin et al., 2016; Huch et al., 2017). However, tissue types derived from organoids may be immature (Chambers et al., 2013; Aurora and Spence, 2016) and limited in thickness and scale because of the absence of abundant vasculature. Additionally, most organoid models can only generate a single

CelPress

Cell Resource

or few developmental lineages (Yin et al., 2016; Jabaudon and Lancaster, 2018; Sato et al., 2009, 2011; Jung et al., 2011; Yin et al., 2016). In this regard, gastruloids, which model early ante-roposterior organization, can recapitulate all germ layers, but they are unable to model later stages of development (Moris et al., 2020).

Here, we propose use of teratomas as a potential model for studying human development (Lensch et al., 2007). Teratomas display multi-lineage differentiation to all germ layers, have a vascularized 3D structure and regions of complex tissue-like organization, and are relatively straightforward to implement. Early teratoma research revealed that teratomas derive from pluripotent germ cells that resemble embryonic cells (Stevens, 1962, 1967; Thurlbeck and Scully, 1960; Stevens and Pierce., 1975). hPSC-derived teratomas are generated by directly injecting hPSCs into immunodeficient mice, where the cells attach and differentiate in a semi-random fashion into all three germ layers (Willis, 1934, 1936; Thurlbeck and Scully, 1960; Böcker, 2002). In this regard, teratoma formation is the gold standard to validate pluripotency and developmental potential of hPSC lines (Smith et al., 2009; Avior et al., 2015).

Leveraging this, there has also been some progress in utilizing the inherent differentiation potential of teratomas to derive highly sought after cell types. For instance, teratomas have been used recently to derive skeletal myogenic progenitors by injecting hPSCs into the *tibialis anterior* muscle of mice to enrich for muscle cell types in the teratomas that formed in those muscles (Chan et al., 2018). Additionally, some groups have successfully enriched for hematopoietic stem cells (HSCs) from teratomas by using strategies such as human umbilical vein endothelial cell (HUVEC) pooling (Suzuki et al., 2013; Tsukada et al., 2017; Philipp et al., 2018; Amabile et al., 2019). However, the semi-random nature of teratoma development has previously made characterization of teratomas difficult, especially as different lineages can often be found in close spatial proximity.

We hypothesized that the advent of high-throughput singlecell gene expression profiling via droplet-based methods (Klein et al., 2015; Macosko et al., 2015; Cao et al., 2017; Rosenberg et al., 2017, 2018; Zheng et al., 2017; Ding et al., 2019) and simple genetic engineering toolsets such as CRISPR-Cas9 could enable us to address this challenge by enabling systematic analysis and perturbation of teratomas at the single-cell level (Qi et al., 2013; Adamson et al., 2016; Black et al., 2016; Dixit et al., 2013; Adamson et al., 2016; Black et al., 2016; Dixit et al., 2016; Chen and Qi, 2017; Datlinger et al., 2017; Akcakaya et al., 2018; van Dijk et al., 2018). Coupled with histology and RNA *in situ* hybridization, we established a comprehensive experimental and computational framework to systematically analyze, perturb, and modulate hPSC-derived teratomas to evaluate their potential for modeling human development and lineage engineering.

RESULTS

Teratoma Characterization

We first characterized teratomas to better understand their growth kinetics, constituent cell types, and spatial organization. We generated 7 teratomas using H1 ESCs, identified cell types using single-cell RNA sequencing (RNA-seq), and validated

these cell types and assessed their spatial organization with histology and RNA fluorescence in situ hybridization (FISH). To generate a teratoma, we made a subcutaneous injection of 5-10 million hESCs into Rag2^{-/-}; $\gamma c^{-/-}$ immunodeficient mice (Figure 1A; STAR Methods). Kinetic trajectories show that it takes an average of around 37 days until we can begin to outwardly see and measure tumor size. We grew the teratomas for up to 70 days until the tumors were of a sufficient size for extraction and downstream analyses (~820 mm²; Figure 1B). Post-extraction, tumors were weighed, inspected, and sectioned (Figure 1C; STAR Methods). We used histology to validate the presence of all 3 germ layers (ectoderm, mesoderm, and endoderm) to confirm pluripotency (Figure 1D; STAR Methods). An independent histology analysis also revealed structures such as developing airways, retinal pigment epithelium and neurons, fetal cartilage and bone, muscle, vasculature, gastro-intestinal (GI) tract, connective tissue, adipocytes, and neuroectoderm (Figure S1A). The remaining tissue was dissociated for single-cell RNA-seq with the droplet-based 10X Genomics Chromium platform (Zheng et al., 2017).

To analyze the resulting sequencing data, we generated single-cell gene expression matrices across the 7 teratomas for human and mouse cells using the CellRanger (Zheng et al., 2017) pipeline from 10X Genomics (STAR Methods; Figure 1A; Table S1A). We removed any teratoma-specific batch effects by using the Seurat data integration pipeline (Stuart et al., 2019) and then clustered the cells using Louvain clustering (Houle et al., 2010). We generated a rough biological annotation of the clusters using a k-nearest neighbors classifier trained on the Mouse Cell Atlas and manually refined the cluster annotations using canonical cell type markers (Han et al., 2018b; Stuart et al., 2019; Tables S2A-S2E). We sub-clustered a cell type expressing ciliated epithelial markers with divergent expression of airway and retinal markers and identified airway epithelium, retinal epithelium, and erythrocytes (Table S2F). We then visualized human and mouse cells with a uniform manifold approximation and projection (UMAP) (Becht et al., 2018) scatterplot (Figure 1E). In human cells, we identified 23 putative cell types across all three germ layers, including endodermal cell types (gut epithelium), ectodermal cell types (early neurons), and an abundance of mesoderm-like cell types that expressed mesenchymal stem cell (MSC)/fibroblast markers, most notably the canonical MSC marker THY1 (An et al., 2018; Figure 1E; Figure S1B; Tables S3A and S3B). We annotated these putative MSC/fibroblast (Fib) cell types as adipogenic (ITM2A, SHOX2), chondrogenic (COL2A1, SOX9), myofibroblasts (COL15A1), or cycling (HMGB2) (Table 1; Table S3C). We visualized the expression of canonical marker genes for each cell type to assess the robustness of our preliminary cell type annotations (Table 1; Figure S1C; Table S3C; STAR Methods).

We further validated the cell type annotations by correlating the expression of each teratoma cell type with the expression of cell types from the Mouse Organogenesis Cell Atlas (Cao et al., 2019), demonstrating that each teratoma cell type generally correlates with at least one fetal mouse cell type (Figure S1D). Although most of the teratoma cell types correlate with the expected mouse cell type, there are some discrepancies that may be due to differences in developmental stage, mouse/



Figure 1. Comprehensive Teratoma Characterization

(A) Schematic of the general workflow. Subcutaneous injection of H1 hPSCs in a slurry of Matrigel and ESC medium was made in the right flank of $Rag2^{-/-};\gamma c^{-/-}$ immunodeficient mice. Weekly monitoring of teratoma growth was quantified by approximating elliptical area (square millimeters). Tumors were extracted after 8– 10 weeks of growth and observed for external heterogeneity before small sections were frozen for H&E staining. The remaining tumor was dissociated into a single-cell suspension via a standard GentleMACS (magnetic cell separation) protocols. A single-cell suspension was used for scRNA-seq (10X Genomics). (B) Growth kinetics of four H1 teratomas.

(C) Images of four teratomas generated from H1 cells.

(D) H&E stains of the four teratoma histology sections. The presence of ectoderm, mesoderm, and endoderm confirmed pluripotency and developmental potential.

(E) UMAP visualization of cell types identified from scRNA-seq of the seven H1 teratomas. Dotted lines separate the cell types originating from each of the 3 germ layers.

See also Figures S1.

CellPress

Cell Resource

Table 1	. Summary of Cell 1	Type Validations					
Germ Layer	Broad Cell Type (Used for CRISPR Screen and miRNA Analysis)	Cell Type	Cells (H1 and Cell Line Teratomas)	Minimal Marker Set	RNA FISH Marker Validation	Identified in Histology Analysis	Mapped to Fetal Human Data
Ecto	neural prog	radial glia	2,579	SOX2, HES5	HES5		yes
		CycProg (cycling neural prog)	1,619	SOX2, HMGB2			yes
	neurons	early neurons	6,010	DCX, MAP2	DCX		yes
		retinal neurons	493	OTX2, NRL		yes	
	retinal epi	retinal epi	7,238	OTX2, MITF, FOXJ1		yes	
	Schwann cell prog (SCP)	Schwann cells	174	MPZ			
		melanoblasts	200	MITF, SOX10, MLANA			
Endo	foregut epi	foregut epi	584	ELF3, PAX9, KRT4			yes
		airway epi	76	FOXJ1, CDHR3	FOXJ1	yes	
	mid/hindgut epi	mid/hindgut epi	1,742	ELF3, CDX2	CDX2		yes
Meso	hematopoietic	immune	1,490	CD74			
		HSC	140	CD34, HHEX			
		erythrocyte	834	GYPA			
	MSC/Fib	adipogenic MSC/Fib	6,487	THY1, ITM2A, SHOX2		yes	
		chondrogenic MSC/Fib	587	THY1, COL2A1, SOX9			
		MSC/Fib	8,046	THY1, COL14A1	THY1		
		cycling MSC/Fib	4,010	THY1, HMGB2		yes	
		MyoFib	3,329	THY1, COL15A1			
	muscle	muscle prog	1,276	MYOD1, PAX7		yes	
		cardiac/skeletal muscle	528	MYOD1, TNNI1, TNNT2	TNNT2	yes	
	pericytes	pericytes	1,053	FOXC1, CYP1B1			
	smooth muscle	smooth muscle	550	ACTA2, RGS5		yes	
		kidney prog	153	WT1			

human-specific expression, as well as the fact that a broad correlation analysis may not be able to distinguish closely related cell types (Figure S1D). For example, HSCs from the teratoma correlate with fetal mouse endothelial cells, potentially reflecting their similar developmental origins (Zovein et al., 2008). The MSC/Fib subtypes as well as pericytes broadly correlate with the same block of mesenchymal fetal mouse cell types, which also reflects their similar developmental origins (Cathery et al., 2018). Retinal pigment epithelia are a type of ependymal cell and thus correlate accordingly (Wolburg et al., 2009). Melanoblasts and retinal neurons are also derived from the neural crest and may share some marker genes, such as MITF, although they are not as closely related as the other cell type correlations discussed previously (Goding, 2000; Mort et al., 2015). Finally, kidney progenitors do not correlate well with any fetal mouse cell type, although there were no kidney cell types in the fetal mouse data at the level of annotation we used (Figure S1D).

Overall, we used canonical marker genes and mouse cell atlases to generate a preliminary annotation of the cell types found in the teratoma single-cell RNA-seq (scRNA-seq) datasets. We provide a summary table of the key marker genes and the experimental and computational validations performed on each cell type in Table 1. In mouse cells, we primarily observed infiltrating immune cells, endothelial cells, and stromal cells (Figure S1E).

Assaying Teratoma Heterogeneity

Assessing heterogeneity between teratomas (especially between teratomas generated from different stem cell lines) is critical for assessing the reproducibility and utility of this model. We generated additional teratomas (Figure 1A) with H9 ESCs, HUES62 ESCs, and PGP1 iPSCs and assessed the cell type composition of the teratomas (Figure 2A; Table S1B). We ran 10X sequencing on each teratoma, integrated the expression profiles, classified cell types using the H1 teratomas as a reference, and visualized the cell types with a UMAP scatterplot (Figure 2B) while also showing the relative contribution of each cell line teratoma to the UMAP embedding (Figure S2A). We also assessed the distribution of cell types represented in each individual H1 teratoma alongside the H9, HUES62, and PGP1 teratomas (Figure 2C; Figure S2B). We then compared the germ layer representation between all teratomas using zebrafish and Mouse Organogenesis Cell Atlas single-cell datasets as references (Wagner et al., 2018; Pijuan-Sala et al., 2019; Figure 2D).



Figure 2. Assaying Teratoma Heterogeneity

(A) Schematic portraying generation of teratomas from multiple cell lines and process for identifying how lines contribute to cell types.

(B) UMAP scatterplot of all cell types present across 3 hPSC lines (H9, HUES62, and PGP1).

(C) Distribution of cell types represented in each individual H1 teratoma, as well as the H9, HUES62, and PGP1 teratomas.

(D) Distribution of germ layer representation in each individual teratoma (along with zebrafish and mouse comparison).

(legend continued on next page)

CelPress

Teratomas are comprised mostly of mesoderm and neuroectoderm, with less endoderm (Figure 2D). The mesoderm is primarily from MSCs/Fibs in H1 teratomas, whereas teratomas from different cell lines show more variability in terms of the MSC/ Fib fraction (Figure 2D; Figure S1B). The relatively low fraction of endoderm in the teratomas as well as the zebrafish and mouse embryo models indicates that endoderm is less prevalent during early development (Figure 2D). Qualitatively, although there is variability in cell type representation among the different teratomas, every teratoma contains most of the major cell types (Figure 2C). By computing the scaled mutual information between cell type assignments and teratoma assignments, we can compute a quantitative metric of this heterogeneity across teratomas (Figure 2E; Kim et al., 2016). We find that the cell type heterogeneity across the H1 teratomas is similar to that of patterned brain organoids (Velasco et al., 2019), whereas the teratomas generated from different cell lines have a much higher level of heterogeneity (Figure 2E). Interestingly, line-specific kinetics were present in regard to teratoma growth, with PGP1 teratomas growing the fastest and HUES62 the slowest (Figure S2C). Some of this accelerated growth may be due to chromosomal abnormalities, because karyotyping has shown that the PGP1 line has material translocated to 7q34 (BRAF) (Figure S2D).

Another key question in teratoma formation is how many cells engraft after stem cell injection. To determine this, for 3 of the 7 H1 ESC teratomas, prior to hPSC injection, cells were transduced with an integrating lentiviral open reading frame (ORF) barcode that can be detected by scRNA-seq (Guo et al., 2018; Figure 2F; Figure S2E). With this barcoding scheme, cells can be labeled individually prior to teratoma formation, and their descendants can be captured after formation via scRNA-seq. Transduced hPSCs were split evenly: half for teratoma formation, and half were frozen down for DNA sequencing. By comparing unique barcodes extracted from genomic DNA in these two cell populations, we can calculate the proportion of cells that engraft. The results showed that, across the three teratomas, over 25% of cells engraft of a total of 10 million injected cells, which suggests that no major bottlenecking occurs during teratoma formation (Figure S2F).

We next tracked barcodes in individual cells by amplifying the expressed barcode from the scRNA-seq library. Because cells from the teratoma with the same barcode originated from the same hPSC, we were able to track whether certain hPSCs were primed to develop into certain lineages. For each cell type, we computed a barcode bias score, which reflects the level to which barcodes tend to be enriched or depleted in that cell type, and plotted this barcode bias along with the total number of barcodes detected in each cell type (Figure 2G; STAR Methods). For each cell type, we also computed a teratoma

Cell Resource

bias score, which reflects how much the proportion of that cell type varies across teratomas, and plotted the correlation of the teratoma bias score with the barcode bias score (Figure 2H; STAR Methods). We found that retinal epithelium is an outlier with a high teratoma bias and a high barcode bias (Figure 2H). Myofibroblast cells also have a relatively high barcode and teratoma bias score, whereas early neurons, radial glia, and mid/ hindgut have a high teratoma bias scores are scaled by the number of cells for each cell type (STAR Methods).

In summary, we found that teratomas generally contain the same major cell types at 10 weeks of growth: a large fraction of MSC/Fib and neuronal cell types and a small fraction of endoderm. Retinal pigment epithelium (RPE) shows a high degree of variability across teratomas and a high level of lineage priming. Notably, the level of heterogeneity between teratomas generated from H1 stem cells is comparable with that observed in organoids (de Souza, 2017; Quadrato et al., 2017; Velasco et al., 2019), but there is a much higher level of heterogeneity among teratomas derived from different hPSC lines. This reflects known epigenetic variability across those lines (Ortmann and Vallier, 2017). Taken together, cellular barcoding confirmed that injected stem cells robustly engraft and contribute to all lineages. This is especially important in the context of using teratomas in highthroughput genetic screens, because one must ensure that there are enough cells contributing to the final tumor so that elements of the genetic screen are not lost due to undersampling, and that majority of the cells retain developmental potential.

Assaying Teratoma Maturity

We next assessed the transcriptional similarity of the teratoma cell types to human fetal cell types using published scRNA-seq datasets from the human neuroectoderm and gut to determine their utility as a tool for modeling human development. We looked at which human embryonic stage the 10-week teratoma cell types most resemble, projected the teratoma data onto the fetal data to assess global transcriptional similarity, and compared the expression of key cell type marker genes (Figure 3A).

Because of the semi-random nature of teratoma differentiation, it is possible that different cell types resemble different stages of embryonic development. Thus, we analyzed individual tissue types separately, looking specifically at the teratoma neuro-ectoderm and gut cell types in depth. We first sub-clustered the neuro-ectoderm cells and identified additional subtypes, including a cluster of early interneurons (Figure 3C; Table S2G). We then compared the average expression of all cells belonging to neural subtypes with the average expression of the same subtypes in a 2,300-cell fetal brain dataset at different

(G) Number of unique barcodes detected in each cell type plotted along with the cell type bias for specific barcodes (computed using the Kullback-Liebler [KL] divergence of cell type identities, with barcode identities scaled by the number of cells in each cell type). Cell types are colored by germ layer. (H) Teratoma bias for each cell type plotted against barcode bias. Cell types are colored by germ layer.

See also Figure S2.

⁽E) Normalized entropy represents how well cell type assignments are mixed with teratoma/organoid/cell line identities. A higher normalized entropy implies less cell type variation between teratomas/organoids/cell lines. The cell line teratomas include one teratoma each generated from the HUES62, H9, and PGP1 lines. (F) H1 cells were uniquely barcoded with lentiviral vectors at low MOI before teratoma formation. The barcodes were counted and assessed for lineage/cell type priming of cells.



(legend on next page)

CelPress

stages of development (Zhong et al., 2018; Figures 3A and 3B). We found that teratoma neuronal cells had scores highly similar to the human prefrontal cortex at gestational weeks 13–17 with the highest score for weeks 16–17 (Figure 3B). Because of the high similarity to week 16–17 human data, we identified the teratoma subtypes (radial glia, cycling progenitors, early neurons, and early interneurons) that matched with the cell types seen in a larger, 40,000+-cell, week 17–18 dataset from the human prefrontal cortex for further analysis (Polioudakis et al., 2019; Figures 3A and 3C).

We then generated similarity weighted non-negative embedding (SWNE) of week 17–18 human prefrontal cortex cells and projected teratoma cells from the matching subtypes onto the fetal human SWNE (Figures 3A and 3D; Wu et al., 2018). We found that similar cell types map to similar spatial positions in the SWNE, suggesting overall similar expression patterns, although the teratoma SWNE shows some overlap between cycling progenitors and radial glia as well as early interneurons and excitatory neurons (Figure 3D). Additionally, the teratoma radial glia cells project onto the fetal intermediate progenitors (Figure 3D).

To further assess the similarity of the teratoma neuro-ectoderm cell types to the fetal prefrontal cortex cell types, we defined a panel of neuronal cell type marker genes (DCX, NEU-ROD1, HES5, SOX2, HMGB2, VIM, and DLX1) and then correlated the expression of these marker genes between the teratoma cells and fetal brain cells for every matched cell type (Figures 3A and 3E). We found a fairly high correlation overall, with R = 0.82 for radial glia, R = 0.93 for cycling progenitors, R = 0.84 for interneurons, and R = 0.77 for early neurons (Figure 3E). We also looked at the cell type proportions in the fetal prefrontal cortex versus the teratoma, showing that the teratoma has far more progenitor cells, such as radial glia, and fewer early neurons with no detectable mature neurons (Figure 3F). We also ran a differential expression as well as a gene set enrichment analysis between the matched teratoma and fetal prefrontal cortex cell types to assess the differences between the teratoma and fetal cells (Figures S3A and S3B). All four cell types showed similar top differentially expressed genes as well as gene sets, suggesting that the main differences between the teratoma and fetal cells are global and not cell type specific (Figure S3A; Figure 3B). The teratoma cells have a higher expression of genes related to organ morphogenesis, whereas the fetal cells express genes related to methylation, suggesting that the teratoma cells

Cell Resource

may not have the same epigenetic signatures as fetal cells (Figures S3A and S3B).

This analysis was repeated with teratoma gut subtypes using a published fetal gut dataset as a reference (Gao et al., 2018). The teratoma gut cells were most similar to gestational week 8–11 gut age (Figure S3C). We compared marker genes for gut cell types (CDX1, CDX2, HHEX, FOXJ1, PAX9, and SOX2) between teratoma and fetal cells and found a high overall correlation, with R = 0.98 for foregut and R = 0.98 for mid/hindgut (Figure S3D). Projecting fetal gut data onto the teratoma SWNE again resulted in relatively similar spatial positioning (Figure S3E). We see that the teratoma produces less foregut and more mid/hindgut than the fetal gut (Figure S3F). When looking at the differences between the teratoma and fetal gut cells, we again see that the fetal cells express more methylation-related genes (Figure S3G and S3H). In this case, the teratoma cells express more genes related to RNA/DNA metabolism (Figure S3G and S3H).

To further validate these results, we used RNAScope in situ hybridization (ISH) to probe for the radial glia marker HES5 and the early excitatory neuron marker DCX, which both showed high abundance in regions of neuro-ectoderm in fixed teratoma tissue sections (Figure 3G). Probes for the genes POLR2A, PPIB, and UBC were used as positive controls and the bacterial marker DapB as a negative control (Figure 3H). Additionally, we probed for FOXJ1 (cilia), CDX2 (intestine epithelium), TNNT2 (cardiac), and THY1 (mesenchyme/Fib) in ciliated airway epithelium, intestinal villi, developing cardiac muscle, and mesenchyme, respectively (Figure S3I). We were able to visualize a high abundance of the respective RNA transcripts and confirm the identity of the respective tissue using H&E staining and histology (Figure S3I). Overall, we were able to show that the teratoma neuro-ectoderm and gut cell types are transcriptionally similar to their fetal counterparts while also identifying the developmental stage of the teratoma cells. We validated the presence of six cell types (2 per germ layer) using RNAScope ISH and histology, which also showed that these cell types contain some degree of spatial organization (Figure 3G; Figure S3I; Table 1). Thus, we were able to further validate the teratoma neuro-ectoderm and gut cell types by mapping them onto reference fetal human scRNA-seq datasets and probing the spatial expression of the canonical marker genes DCX, HES5, and CDX2 (Table 1; Table S3C). We also probed the spatial expression of FOXJ1, TNNT2, and THY1, adding more evidence to the ciliated epithelium, cardiac muscle, and MSC/Fib cell type annotations (Table 1; Table S3C).

Figure 3. Assaying Teratoma Maturity

⁽A) Teratoma neuro-ectoderm cell types were mapped to fetal cortical cell types, and the corresponding teratoma cell types were projected onto SWNEs of fetal cells. Key marker genes were correlated across matching teratoma/fetal cell types, and average expression of teratoma cell types was correlated with fetal cell types from different stages of development.

⁽B) Cosine similarity of teratoma brain cells with fetal brain cells of different ages.

⁽C) UMAP embedding of teratoma neuro-ectoderm sub-clusters (Table S2G).

⁽D) Projection of teratoma neuro-ectoderm cell types onto the SWNE of fetal cortical cells.

⁽E) Correlation of the scaled expression of key marker genes across radial glia, cycling progenitors, early neurons, and interneurons. Marker genes are colored by their corresponding cell type.

⁽F) Fraction of brain-related cell types in the teratoma and fetal cortex.

⁽G) H&E stain (left) and RNAScope image (right) of HES5 (radial glia marker, top) and DCX (early neuron, bottom) expression. DAPI is a nuclear stain. 4–10 punctate dots/cell is a positive result. Dots were dilated using ImageJ. Scale bars, 50 μM.

⁽H) Positive (top) and negative (bottom) RNAScope control staining. DAPI is a nuclear stain. 4–10 punctate dots/cell is a positive result. Scale bars, 50 µM. See also Figure S3.



CellPress



(legend on next page)

CellPress

Engineering Teratomas via Genetic Perturbations

To establish the utility of the teratoma system as a model for human development, we next performed a single-cell genetic knockout screen using CRISPR-Cas9. Specifically, we focused on 24 major organ/lineage specification genes that are embryonic lethal upon knockout in mice (Table S4A). Studying the effects of these genes using cell lines or organoid models would typically require different experiments and different models for each cell lineage because even a single gene can have functions across multiple cell types and even different germ layers. With the teratoma model, we can, in principle, screen the effects of these genetic perturbations in all major cell lineages and germ layers in the same experiment. Using the CRISPR droplet sequencing (CROPseq) Guide-Puro vector backbone, we cloned in 48 individual single guide RNAs (sgRNAs) directed at each developmental gene (2 sgRNAs per gene) (Datlinger et al., 2017; Figure 4A; Table S4B). We also designed a stable Cas9-expressing iPSC line (PGP1) to prevent Cas9 silencing (Figures S4A and S4B; STAR Methods). After creating a pooled lentivirus library with our sgRNAs, we transduced our engineered PGP1-Cas9 line at a MOI of 0.1 so that each cell received approximately one perturbation (Figure 4A). After selection, these cells were injected subcutaneously into 3 Rag2^{-/-}; $\gamma c^{-/-}$ immunodeficient mice for teratoma formation, extraction, and downstream scRNA-seq processing with 10X Genomics (Figure 4A).

We validated the editing efficiencies of all of our guide RNAs using PCR amplification of the expected cut site and looked for mutations and insertions or deletions (indels) with CRISP-Resso (Tables S4C and S4D; STAR Methods). We then selected the top guide targeting each gene, which resulted in a total of 16 guides (Tables S1C, S4C, and S4D; STAR Methods). We then only used these validated guides for further computational analysis. To assess the reproducibility of our results, we also reran the CRISPR knockout (KO) screen by repooling these validated guides and generated 3 additional teratomas (Figure 4A; Table S1D; STAR Methods). We successfully captured a median of

Cell Resource

118 cells per gene/guide in the original screen and 1,280 cells per gene/guide in the replicate screen (Figure S4C). We were able to capture more cells per guide in the replicate screen because we only pooled the top 16 guides, whereas the original screen had a total of 48 guides (STAR Methods).

To ensure consistent cell types across teratomas, we integrated all six teratomas across the original and replicate screen using Seurat v.3 (Stuart et al., 2019). We then called cell types in the PGP1 teratoma cells using Seurat label transfer with the 7 H1 teratomas as a reference and collapsed developmentally similar cell types (Figure S4D; STAR Methods). To determine the total effect of each KO, we measured the difference in cell type composition between cells in each gene KO with all cells belonging to the non-targeting control (NTC) separately for each screen using earth mover's distance (EMD) (Chen et al., 2020; Figure 4A; STAR Methods). For the original and replicate screen, we ran a ridge regression model to assess the effects of each gene KO on cell type enrichment/depletion (Dixit et al., 2016; Figure 4A; STAR Methods). For each gene, we plotted its EMD along with the Pearson correlation of the regression coefficients for the original screen and the replicate screen, giving us a sense of the effect size and reproducibility of each gene KO (Figure 4B; STAR Methods). We also see that gene KOs with strong effect sizes tend to be more reproducible (R = 0.59) (Figure 4B; STAR Methods). We highlighted genes with a Pearson correlation of greater than 0.4 between the original and replicate screen for further analysis (Figure 4B).

For the highlighted genes *TWIST1*, *RUNX1*, *CDX2*, *KLF6*, and *ASCL1*, we wanted to identify the gene KO effects on cell types that were statistically significant. We merged the cells from both screens and ran a combined ridge regression analysis, computing p values using a permutation test and false discovery rates using the Benjamini-Hochberg correction (STAR Methods). We then visualized all gene KO effects with a false discovery rate (FDR) of less than 0.1 (Figure 4C; STAR Methods).

CDX2 is known to be important for development of the midgut and hindgut (Silberg et al., 2002; Gao et al., 2009a). Our data

Figure 4. Engineering Teratomas via Genetic Perturbations and miRNA-Based Molecular Sculpting

For a Figure360 author presentation of this figure, see https://doi.org/10.1016/j.cell.2020.10.018.

(A) PGP1-Cas9 iPSCs were transduced with a CRISPR library targeting a panel of 16 key developmental genes with 1 sgRNA per gene. After generating 3 teratomas with the PGP1-iPSCs, scRNA-seq was used to identify shifts in cell type formation as a result of gene KOs. We repeated this process with 3 additional teratomas to serve as a replicate screen.

(B) Average effect of gene KO on cell type enrichment/depletion versus the correlation of cell type enrichment between the original screen and replicate screen. Genes with a reproducibility greater than 0.4 (STAR Methods) were selected for further analysis and are highlighted in red. NTCs are highlighted in blue.

(C) A heatmap of the effect size (regression coefficient) of gene KO enrichment for cell types and germ layers. The effect size represents to the level of cell type enrichment (red) or depletion (blue).

(D) Scatterplot of individual guide RNA effects on cell type abundance for the selected genes TWIST1, RUNX1, CDX2, KLF6, and ASCL1. Cell types are colored by their germ layer.

(E) Schematic of the miRNA-HSV-tk-GFP construct. 2A encodes for a self-cleaving peptide. Upon transcription, the expression is diminished when corresponding endogenously expressed miRNA is present in the cell.

(F) Schematic of how a developing teratoma forms in the presence of ganciclovir (GCV; 80 mg/kg/day; STAR Methods) when cells were transduced with a neural-specific miRNA-HSV-tk construct.

(G) Quantification using flow cytometry and gating based on the presence or absence of GFP in 35-day self-patterned whole-brain organoid single cells transduced with HSV-tk-GFP control (red) or miR-124-HSV-tk-GFP (blue).

(H) *In vivo* studies of miR-124-HSV-tk-GFP teratomas in the presence of GCV administration (80 mg/kg/day; STAR Methods) using intratumoral (IT, red) and both intratumoral and intraperitoneal (IPIT, teal) injection methods. A heatmap shows cell type fraction log fold change for each teratoma replicate compared with a control miR-124-HSV-tk-GFP teratoma in the absence of GCV. *Z* scores for each cell type fraction change are plotted as well, with standard deviations calculated using a pooled variance (STAR Methods).

See also Figures S4 and S5.

Figure360⊳



show that cells with a CDX2 are enriched in the foregut and depleted in the mid/hindgut, which lines up with past literature reports that CDX2 KO shifts the gut differentiation pathway away from intestine and toward gastric activation (Simmini et al., 2014; Kim and Shivdasani, 2016; Figures 4C and 4D). TWIST1 showed the largest effect size and is a known transcription factor for epithelial-to-mesenchymal transition (EMT), which is important in development as well as metastatic cancers (Figure 4B; Yang et al., 2004; Kalluri and Weinberg, 2009). Our screen found that cells with TWIST1 KO are depleted in mesodermal cell types (muscle, smooth muscle, pericytes, and MSCs/Fibs) and enriched in neuro-epithelium (retinal epithelium and neurons), confirming prior studies that have identified TWIST1 as key for mesodermal specification (Qin et al., 2012; Figures 4C and 4D). We see that RUNX1 KO results in depletion of neurons and muscle cell types and enrichment in mid/hindgut, which is consistent with previous mouse and stem cell studies that show RUNX1 to be critical for neural crest formation, signaling in gut epithelium stem cells, and myoblast proliferation (Marmigère et al., 2006; Fijneman et al., 2012; Scheitz and Tumbar, 2013; Umansky et al., 2015; Sarper et al., 2018; Figures 4C and 4D). KLF6 KO resulted in depletion of pericytes, consistent with its role in promoting endothelial activation during vascular repair (Garrido-Martín et al., 2013; Figures 4C and 4D). Surprisingly, the ASCL1 KO resulted in an increase in the proportion of retinal epithelium and neural progenitors (Figures 4C and 4D). Because ASCL1 is key for cell cycle exit and neuronal differentiation, knocking out ASCL1 may slow down neurogenesis and result in a buildup of neural progenitors (Castro et al., 2011). With this CRISPR KO screen of key developmental requlators, we were able to simultaneously assay the multi-lineage functions of these genes in a human-specific model, something that, to our knowledge, other human developmental models cannot currently readily accomplish.

Modeling Neural Disorders Using Teratomas

Although we were able to demonstrate the teratomas' unique ability to assess the multi-lineage function of embryonic lethal genes, we also wanted to see whether teratomas could model human neural disorders. Specifically, we looked into Pitt-Hopkins (Dean, 2012), Rett (Ehinger et al., 2018), and L1 (Stumpel and Vos, 1993) syndromes. Pitt-Hopkins syndrome is a rare neurodevelopmental disorder most often caused by de novo loss of function of one allele of the transcription factor 4 (TCF4) gene (Forrest et al., 2014). Rett syndrome is a severe X-linked neurological disorder caused by a de novo mutation in the methyl-CpG-binding protein 2 (MECP2) gene. Finally, L1 syndrome is another X-linked syndrome with a mutation in the L1 cell adhesion molecule (L1CAM) gene, important for neuron migration, adhesion, and neuronal differentiation (Samatov et al., 2016). To assess the downstream effects of perturbing these genes, we generated a CRISPR KO library targeting TCF4, MECP2, and L1CAM, with 3 guides for each gene (Table S5A). We transduced PGP1-Cas9 cells with the neural disorder library, generated 2 teratomas, and then sequenced 2 scRNA-seq libraries for each teratoma using the 10X Genomics platform (Table S1E; Zheng et al., 2017).

We integrated and clustered the teratomas using Seurat data integration and used Seurat's label transfer method to call cell



types using the H1 teratomas as the reference. We then looked for shifts in cell type proportion and cell-type-specific gene expression as a result of the gene KOs (Figure S4E). As expected, we found that the shift in cell type proportion (normalized EMD) was much smaller than for the embryonic lethal KOs (Figure S4F). We thus looked at cell-type-specific shifts in gene expression from the neurological disorder KOs instead. We merged our cell types into 7 broad cell types (neurons, neural progenitors, gut, retinal epithelium, muscle, immune, and MSC/Fib) and computed the differential expression between each gene KO and the NTCs (STAR Methods). There was no significant gene expression shift because of the presence of a double-stranded break (per control guides that target the adenoassociated virus integration sites [AAVSs]) (Table S5B).

We then analyzed the effect of L1CAM in neurons and the effect of TCF4 and MECP2 in neural progenitors and plotted the cell-type-specific log fold changes for all differentially expressed genes (DEGs) with an FDR below 0.1 across both teratomas, showing that our hits are fairly reproducible (Figures S4G-S4I). Knocking out L1CAM in neurons decreased the expression of clusterin (CLU), an effect that has been shown previously in colorectal cancer cells (Shapiro et al., 2015), while increasing the expression of MAPT (which produces the tau protein). Tau efflux via L1CAM exosomes is present in certain neurological diseases (Shi et al., 2016; Figure S4G; Table S5C). Knocking out MECP2 in neural progenitors decreased the expression of transient receptor potential cation channel subfamily M member 3 (TRPM3), and previous literature has shown a similar decrease in expression and function of transient receptor potential (TRP) channels in the hippocampus and several other brain regions of MECP2 mutant mice, contributing to Rett syndrome etiology (Chapleau et al., 2013; Li and Pozzo-Miller, 2014; Suzuki et al., 2016; Figure S4H; Table S5D). Finally, knocking out TCF4 in neural progenitors decreased the expression of FOXO3, which is consistent with TCF4 knockdown studies in the human neuroblastoma line SH-SY5Y showing a fold decrease in FOXO3. which has been suggested to contribute to the molecular pathology of Pitt-Hopkins syndrome and other autism spectrum disorders (Forrest et al., 2013; Figure S4I; Table S5E). Overall, we were able to reproducibly discover cell-type-specific gene expression shifts that occurred when knocking out the genes underlying Rett, Pitt-Hopkins, and L1 syndromes, potentially building a resource for future in-depth study.

Engineering Teratomas via miRNA-Based Molecular Sculpting

Because teratomas are vascularized and have the potential to yield mature tissue, we next sought to sculpt teratomas toward specific lineages, which could allow focused developmental modeling studies and large tissue engineering. We used endogenously expressed microRNAs (miRNAs) (Ambros, 2004; Bartel, 2004; Bartel, 2018), which are often unique to specific cell types, lineages, or disease states (Lu et al., 2005; Shivdasani, 2006). Specifically, we appended tissue-specific miRNA target sequences to the 5' and 3' UTRs of a GFP suicide gene (HSV-tk-GFP), suppressing its expression in a miRNA-specific lineage of interest (Figure 4E; Table S5G; Miki et al., 2015; Nissim et al., 2017; Hirosawa et al., 2017). This design ensures that

CellPress

cell types that do not express the miRNA are killed by the suicide gene in the presence of gancliclovir (GCV), selecting our desired lineage (Figure 4F).

We first tested the functionality our miRNA-HSV-tk-GFP constructs in H1 ESCs by showing that cells transduced with our miRNA-HSV-tk-GFP construct die in the presence of 10 μ M GCV after 5 days of culture, whereas cells transduced with a GFP control continue to grow (Figure S5A). We then assessed the cell type specificity of the miRNA construct using miR-21-expressing HeLa cells (Lu et al., 2008; Medina and Slack, 2008; Yao et al., 2009; Bartel, 2018). HEK293T cells show little to no expression of miR-21 and can serve as a control (Zhu et al., 2008; Li et al., 2009; Chak et al., 2016). After transduction of both cell lines with our miR-21-HSV-tk-GFP construct, we cultured the cells for 5 days and then performed flow cytometry analysis, where we saw a decrease in GFP expression in HeLA cells but not in HEK293T cells (Figure S5B). This indicates that GFP expression was silenced by the miR-21 expressed by HeLa cells. We used an HSV-tk-GFP construct without any miRNA binding sites as a control (Figure S5B). We repeated this experiment with a miR-126-HSV-tk-GFP construct (endothelial cell specific; Wang et al., 2008) and observed a decrease in GFP signal in HU-VECs compared with the HEK293T control (Figure S5C). With these, we were able to validate HSV-tk killing with GCV and the ability of our miRNA constructs to specifically repress GFP in target cell lines.

We next validated our constructs in whole brain organoids. Following a standard self-patterned whole-brain organoid protocol (Figure S5D; STAR Methods; Quadrato et al., 2017), we created organoids using H1 ESCs transduced with the miR-124-HSV-tk-GFP construct or the HSV-tk-GFP construct (lacking any miRNA binding sites). We used miR-124 because it is a pan-neural miRNA (Lagos-Quintana et al., 2002; Smirnova et al., 2005; Sun et al., 2015). Day 35 organoids from both groups (miR-124-HSV-tk-GFP and HSV-tk-GFP) were dissociated down to the single-cell level and analyzed via flow cytometry for GFP fluorescence (STAR Methods). As expected, HSV-tk-GFP organoid single cells maintained their GFP fluorescence, whereas miR-124-HSV-tk-GFP organoids showed loss of GFP expression (Figure 4G).

We then tested the miRNA-HSV-tk-GFP constructs in vivo using the miR-124-HSV-tk construct to generate teratomas enriched for the neural lineage. After the H1 ESC line was successfully transduced with the miR-124-HSV-tk-GFP construct, we formed teratomas as described in our previous studies (STAR Methods). When teratomas reached a minimum of 1 cm in diameter, we began intratumoral (IT) injections with GCV (80 mg/kg/ day; STAR Methods) or two-site intraperitoneal and IT (IPIT) injections (50/mg/kg/day for each site; STAR Methods), all compared with a control miR-124-HSV-tk-GFP teratoma with no GCV (STAR Methods). There were 2 teratomas for each injection condition for a total of 4 teratomas and 1 control teratoma, and all teratomas were grown for up to 70 days. Post-extraction, teratomas were observed for external heterogeneity. Teratomas that received GCV injections were of smaller size (approximately 2 cm compared with 4 cm) and weight (approximately 1-2 g compared with 5+ g) than the control teratoma without GCV injections (Figure S5E).

Cell Resource

We ran the 10X scRNA-seq protocol on each teratoma and classified cells using Seurat label transfer (Table S1F) (Stuart et al., 2019). A comparison of the GCV+ teratoma cell type composition with the GCV-teratoma revealed enrichment in early neurons, neuronal progenitors, and Schwann cells (Figure 4H). In addition, we saw depletion in muscle, RPE (which lacks miR-124 expression), and other cell types (Figure 4H). Teratomas with the IPIT injection strategy showed stronger enrichment for the neuro-ectoderm cell types, suggesting that addition of an intraperitoneal injection site helps with GCV selection (Figure 4H). We also visualized the neuro-ectoderm enrichment in GCV+ teratomas with H&E staining of a GCV+ and GCV- teratoma (Figure S5F). IPIT teratomas had stronger enrichment for early neurons (Z score > 3) than for neuronal progenitors or Schwann cells, possibly because expression of miR-124 increases as the neuro-ectoderm cell types mature (Figure 4H; Figure S5F).

We further validated the enrichment of neuro-ectoderm in IPIT teratomas by immunostaining for *PAX6*, a key marker of neuronal fate determination (Figure S5G). The three GCV+ teratoma sections with IPIT injections showed higher levels of *PAX6* protein expression than the three GCV- teratoma sections, confirming that our miR-124 construct enriches for neuro-ectoderm (Figure S5G). We used a secondary antibody (Dylight 550) to confirm that there was no non-specific secondary antibody binding (Figure S5H). Additionally, we confirmed that the GCV+ teratoma has higher expression of HES5, a key radial glia marker, using RNA FISH (Figure S5I).

In summary, we developed a miRNA circuit that enables us to engineer teratomas toward a desired lineage. Our *in vivo* results showed that administering GCV through multiple sites resulted in improved neuro-ectoderm enrichment. Our miRNA circuit can be extended to any cell-type-specific miRNA and could have applications for studying developmental biology and human disease as well as for tissue engineering.

DISCUSSION

Teratomas have the potential to be a useful multi-lineage model of human development. Their major advantages are that they can grow to a large size because of their vascularization, and they can produce a wide array of cell types from all major developmental lineages. Additionally, as we demonstrated with our CRISPR-Cas9 KO screens, a teratoma's ability to generate cells from all lineages enables simultaneous pan-tissue assessment of the effect of genetic perturbations on human development in a single integrated experiment. Furthermore, we show that teratomas can be engineered using miRNA circuits to grow/enrich specific tissues of interest *in vivo*.

Future studies with this model could explore increasing tissue maturity with extended growth, perhaps with the use of larger animal hosts. Benchmarking with human patient-derived teratomas would also be valuable, especially because many of these can become quite mature. Another critical future study is assessing the effect of different dissociation methods on teratoma cell type proportion. The ability to achieve greater cell numbers with the most current single-cell RNA sequencing protocols, such as split-pool ligation-based transcriptome sequencing (SPLiT-seq) (Rosenberg et al., 2018) and single-cell combinatorial indexing



RNA sequencing (sci-RNA-seq) (Cao et al., 2017), will be vital for identifying additional cell types. A time series analysis of teratomas at multiple stages of maturity could help uncover the developmental pathways the cell types follow. Additionally, pooling different cell types with hPSCs prior to injection may help aid with cellular enrichment/maturity in the teratoma (i.e., HUVECs to enrich for HSC populations; Philipp et al., 2018) or enriching for desired cell types based on injection site (Chan et al., 2018). Growing patient-specific teratomas could benefit disease research through isogenic iPSC lines aiding with understanding the disease state in various tissues that may be inaccessible with current technologies. Finally, further optimization of the miRNA molecular sculpting approach is necessary; specifically, generating stable miRNA circuit-expressing cell lines by insertion in constitutively active loci such as AAVS1 and optimizing the timing, dose, and route of GCV administration. Taken together, we believe that teratomas are a promising platform for modeling multi-lineage human development, pan-tissue functional genetic screening, and tissue engineering.

Limitations of Study

Every model system has its intrinsic strengths and weaknesses, and below we discuss some of the limitations of the teratoma system and also considerations regarding improving it to further enable basic science and engineering studies. One issue with the teratoma system (and organoids) is its intrinsic degree of heterogeneity (de Souza, 2017; Quadrato et al., 2017; Capowski et al., 2019; Phipson et al., 2019). In this regard, we found use of internal controls when conducting perturbation experiments improved the signal in our studies. For example, in our CRISPR-Cas9 screen, each teratoma contained gene targeting guides and NTCs, enabling us to compare cell type proportion shifts within each teratoma without having to worry about heterogeneity between teratomas.

Although teratomas have regions of organization and maturity, these may develop in an asynchronous manner. This lack of synchronization may prove to be a barrier when accessing certain mature cell types that need a highly ordered cellular context to develop.

Also, because teratomas contain cell types from all lineages, finding a single dissociation protocol that captures as many cell types as possible is a challenge. The choice of dissociation method can drastically change the cell types profiled in scRNA-seq, and it is likely that the set of cell types we see in our data is biased by our dissociation protocol (Denisenko et al., 2019). It may be the case that no single dissociation method can capture all cell types, and it will be necessary to design specific dissociation protocols to capture specific tissues.

Additionally, our cell type annotations are still preliminary. Although we validated key cell types by comparison with fetal human/mouse reference datasets and RNA FISH, we were not able to validate all cell types because of limited developmental human reference scRNA-seq datasets, as well as cost constraints. Thus, some cell types, such as neuro-ectoderm cell types, have more validation than others, giving us greater confidence in their identity (Table 1). We may also still be underpowered to detect less abundant cell types, and additional scRNA-



seq could enable us to resolve some missing cell types because undersampling could result in smaller cell types being collapsed into a larger cell type during analysis.

In regard to lineage engineering, we anticipate that there will be a considerable degree of silencing that occurs in the miRNA-suicide gene constructs because of the use of lentiviral vectors. Future studies could explore incorporating these in genomic regions, such as the AAVS1 locus, that would enable constitutive expression across all cell types. Safety switches based on suicide genes will also be critical for eliminating potential residual undifferentiated cells and mouse cells within the teratoma to mitigate effects on safety and utility in tissue engineering applications.

STAR***METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell Culture
 - Organoid Generation and Dissociation
 - Animals
- METHOD DETAILS
 - PGP1-Cas9 Clone Generation
 - o sgRNA Design
 - Library Preparation
 - Viral Production
 - Viral Transduction
 - sgRNA Editing Rate Validation
 - O GCV-HSV-tk Killing in vitro
 - miRNA-HSV-tk-GFP Knockdown in vitro
 - Teratoma Formation
 - Molecular Sculpting of Teratomas
 - Teratoma Processing
 - Histology and RNAScope®
 - Immunostaining
 - Microscopy
 - Cost Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Overview
 - Single Cell RNA-seq Processing
 - Seurat Data Integration
 - H1 Teratoma Clustering and Validation
 - Quantitative Assessment of Teratoma Heterogeneity and Cell Type Bias
 - O Lentiviral Barcode and CRISPR Guide Assignment
 - H1 Cell Barcoding Analysis
 - Developmental Staging Analysis
 - O PGP1 Embryonic Lethal Screen Analysis
 - O PGP1 Neural Disorder Screen Analysis
 - Molecular Sculpting Analysis
 - Figure Generation

CellPress

Cell Resource

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j. cell.2020.10.018.

ACKNOWLEDGMENTS

We thank members of the Mali lab for advice on and help with experiments; Dr. Peter Kharchenko for suggestions on how to perform the computational analysis; Alexander Militar for assistance with schematic generation; Marianna Yusupova for help with initial studies; and the Moore's Cancer Center Histology Core, UC San Diego Microscopy Core, Sanford Consortium Flow Cytometry Core, and IGM Genomics Center for help with sample processing. D.M. thanks Nakon Aroonsakool, in loving memory. This work was generously supported by UCSD institutional funds and NIH grants R01HG009285, R01CA222826, R01GM123313, and U54CA209891.

AUTHOR CONTRIBUTIONS

D.M., Y.W., K.Z., and P.M. conceived and designed the studies and wrote the manuscript with input from all authors. D.M. and J.T. generated teratomas, performed RNAScope studies, and performed scRNA-seq experiments. Y.W. performed quantification and statistical analysis, including mapping and comparisons with published datasets. A.D. generated organoids, and A.D. and U.P. helped with experimental consultation and assistance. D.Z. engineered the PGP1-Cas9 line and designed the developmental screen gene list. M.H. helped with microscopy and kinetic growth measurements. A.T. assessed histology images and identified cell types.

DECLARATION OF INTERESTS

D.M., Y.W., K.Z., and P.M. have filed patents based on this work. K.Z. is a cofounder, equity holder, and paid consultant of Singlera Genomics, which has no commercial interests related to this study. P.M. is a scientific co-founder of Shape Therapeutics, Boundless Biosciences, Seven Therapeutics, Navega Therapeutics, and Engine Biosciences, which have no commercial interests related to this study. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policy.

Received: August 10, 2019 Revised: June 6, 2020 Accepted: October 9, 2020 Published: November 4, 2020

REFERENCES

Abdi, H., and Williams, L.J. (2010). Principal component analysis. WIREs Computational Systems 2, 433–459.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1882.e21.

Akcakaya, P., Bobbin, M.L., Guo, J.A., Malagon-Lopez, J., Clement, K., Garcia, S.P., Fellows, M.D., Porritt, M.J., Firth, M.A., Carreras, A., et al. (2018). In vivo CRISPR-Cas gene editing with no detectable genome-wide off-target mutations. Nature *561*, 416–419.

Amabile, G., Welner, R.S., Nombela-Arrieta, C., D'Alise, A.M., Di Ruscio, A., Ebralidze, A.K., Kraytsberg, Y., Ye, M., Kocher, O., Neuberg, D.S., et al. (2019). In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells. Blood *121*, 1255–1264.

Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350–355. An, Z., Sabalic, M., Bloomquist, R.F., Fowler, T.E., Streelman, T., and Sharpe,

P.T. (2018). A quiescent cell population replenishes mesenchymal stem cells to drive accelerated growth in mouse incisors. Nat. Commun. 9, 378.

Aurora, M., and Spence, J.R. (2016). hPSC-derived lung and intestinal organoids as models of human fetal tissue. Dev. Biol. *420*, 230–238.

Avior, Y., Biancotti, J.C., and Benvenisty, N. (2015). TeratoScore: Assessing the Differentiation Potential of Human Pluripotent Stem Cells by Quantitative Expression Analysis of Teratomas. Stem Cell Reports *4*, 967–974.

Bansod, S., Kageyama, R., and Ohtsuka, T. (2017). Hes5 regulates the transition timing of neurogenesis and gliogenesis in mammalian neocortical development. Development 144, 3156–3167.

Barkas, N., Petukhov, V., Nikolaeva, D., Lozinsky, Y., Demharter, S., Khodosevich, K., and Kharchenko, P.V. (2019). Joint analysis of heterogeneous singlecell RNA-seq dataset collections. Nat. Methods *16*, 695–698.

Bartel, D.P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function Genomics: The miRNA Genes. Cell *116*, 281–297.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell 173, 20-51.

Bartfeld, S., Bayram, T., van de Wetering, M., Huch, M., Begthel, H., Kujala, P., Vries, R., Peters, P.J., and Clevers, H. (2015). In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. Gastroenterology *148*, 126–136.e6.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. Published online December 3, 2018. https://doi.org/10.1038/nbt.4314.

Bigorgne, A.E., Farin, H.F., Lemoine, R., Mahlaoui, N., Lambert, N., Gil, M., Schulz, A., Philippet, P., Schlesser, P., Abrahamsen, T.G., et al. (2014). TTC7A mutations disrupt intestinal epithelial apicobasal polarity. J. Clin. Invest. *124*, 328–337.

Black, J.B., Adler, A.F., Wang, H.G., D'Ippolito, A.M., Hutchinson, H.A., Reddy, T.E., Pitt, G.S., Leong, K.W., and Gersbach, C.A. (2016). Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell *19*, 406–414.

Böcker, W. (2002). WHO classification of breast tumors and tumors of the female genital organs: pathology and genetics. Verh. Dtsch. Ges. Pathol. *86*, 116–119.

Boj, S.F., Hwang, C.I., Baker, L.A., Chio, I.I., Engle, D.D., Corbo, V., Jager, M., Ponz-Sarvise, M., Tiriac, H., Spector, M.S., et al. (2015). Organoid models of human and mouse ductal pancreatic cancer. Cell *160*, 324–338.

Bort, R., Martinez-Barbera, J.P., Beddington, R.S., and Zaret, K.S. (2004). Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. Development *131*, 797–806.

Brown, J., Quadrato, G., and Arlotta, P. (2018). Studying the Brain in a Dish: 3D Cell Culture Models of Human Brain Development and Disease. Curr. Top. Dev. Biol. *129*, 99–122.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive singlecell transcriptional profiling of a multicellular organism. Science *357*, 661–667.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502.

Capowski, E.E., Samimi, K., Mayerl, S.J., Phillips, M.J., Pinilla, I., Howden, S.E., Saha, J., Jansen, A.D., Edwards, K.L., Jager, L.D., et al. (2019). Reproducibility and staging of 3D human retinal organoids across multiple pluripotent stem cell lines. Development *146*, dev171686.

Castro, D.S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L.G., Hunt, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. Genes Dev. *25*, 930–945.

Cathery, W., Faulkner, A., Maselli, D., and Madeddu, P. (2018). Concise Review: The Regenerative Journey of Pericytes Toward Clinical Translation. Stem Cells *36*, 1295–1310.





Chak, K., Roy-Chaudhuri, B., Kim, H.K., Kemp, K.C., Porter, B.E., and Kay, M.A. (2016). Increased precursor microRNA-21 following status epilepticus can compete with mature microRNA-21 to alter translation. Exp. Neurol. *286*, 137–146.

Chambers, S.M., Tchieu, J., and Studer, L. (2013). Build-a-brain. Cell Stem Cell *13*, 377–378.

Chan, S.S.K., Arpke, R.W., Filareto, A., Xie, N., Pappas, M.P., Penaloza, J.S., Perlingeiro, R.C.R., and Kyba, M. (2018). Skeletal Muscle Stem Cells from PSC-Derived Teratomas Have Functional Regenerative Capacity. Cell Stem Cell 23, 74–85.e6.

Chapleau, C.A., Lane, J., Pozzo-Miller, L., and Percy, A.K. (2013). Evaluation of current pharmacological treatment options in the management of Rett syndrome: from the present to future therapeutic alternatives. Curr. Clin. Pharmacol. *8*, 358–369.

Chen, M., and Qi, L.S. (2017). Repurposing CRISPR System for Transcriptional Activation. In RNA Activation, L.-C. Li, ed. (Springer), pp. 147–157.

Chen, W.S., Zivanovic, N., van Dijk, D., Wolf, G., Bodenmiller, B., and Krishnaswamy, S. (2020). Uncovering axes of variation among single-cell cancer specimens. Nat. Methods *17*, 302–310.

Clevers, H. (2016). Modeling Development and Disease with Organoids. Cell 165, 1586–1597.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (Routledge).

Collin, J., Queen, R., Zerti, D., Dorgau, B., Hussain, R., Coxhead, J., Cockell, S., and Lako, M. (2019). Deconstructing Retinal Organoids: Single Cell RNA-Seq Reveals the Cellular Components of Human Pluripotent Stem Cell-Derived Retina. Stem Cells *37*, 593–598.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301.

de Souza, N. (2017). Organoid variability examined. Nat. Methods 14, 655.

Dean, L. (2012). Pitt-Hopkins Syndrome. In Medical Genetics Summaries, V.M. Pratt, H.L. McLeod, W.S. Rubinstein, S.A. Scott, L.C. Dean, B.L. Kattman, and A.J. Malheiro, eds. (National Center for Biotechnology Information).

Dekkers, J.F., Wiegerinck, C.L., de Jonge, H.R., Bronsveld, I., Janssens, H.M., de Winter-de Groot, K.M., Brandsma, A.M., de Jong, N.W., Bijvelds, M.J., Scholte, B.J., et al. (2013). A functional CFTR assay using primary cystic fibrosis intestinal organoids. Nat. Med. *19*, 939–945.

Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R.K., Lister, R., and Forrest, A.R.R. (2019). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol. *21*, 130.

Ding, J., Zhang, Y., Zhang, L., Li, Z., Fang, Q., Gao, R., and Zhang, Z. (2019). Systematic comparative analysis of single cell RNA-sequencing methods. Genome Biol. *20*, 242.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1866.e17.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dutta, D., Heo, I., and Clevers, H. (2017). Disease Modeling in Stem Cell-Derived 3D Organoid Systems. Trends Mol. Med. *23*, 393–410.

Ehinger, Y., Matagne, V., Villard, L., and Roux, J.C. (2018). Rett syndrome from bench to bedside: recent advances. F1000Res. 7, 398.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360*, eaar3131.

Fijneman, R.J.A., Anderson, R.A., Richards, E., Liu, J., Tijssen, M., Meijer, G.A., Anderson, J., Rod, A., O'Sullivan, M.G., Scott, P.M., and Cormier, R.T.

(2012). Runx1 is a tumor suppressor gene in the mouse gastrointestinal tract. Cancer Sci. *103*, 593–599.

Fligor, C.M., Langer, K.B., Sridhar, A., Ren, Y., Shields, P.K., Edler, M.C., Ohlemacher, S.K., Sluch, V.M., Zack, D.J., Zhang, C., et al. (2018). Three-Dimensional Retinal Organoids Facilitate the Investigation of Retinal Ganglion Cell Development, Organization and Neurite Outgrowth from Human Pluripotent Stem Cells. Sci. Rep. 8, 14520.

Forrest, M.P., Waite, A.J., Martin-Rendon, E., and Blake, D.J. (2013). Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. PLoS ONE 8, e73169.

Forrest, M.P., Hill, M.J., Quantock, A.J., Martin-Rendon, E., and Blake, D.J. (2014). The emerging roles of TCF4 in disease and development. Trends Mol. Med. *20*, 322–331.

Fougerousse, F., Bullen, P., Herasse, M., Lindsay, S., Richard, I., Wilson, D., Suel, L., Durand, M., Robson, S., Abitbol, M., et al. (2000). Human-mouse differences in the embryonic expression patterns of developmental control genes and disease genes. Hum. Mol. Genet. 9, 165–173.

Friedman, A.J., Hastie, T., Tibshirani, R., Balasubramanian, N., Tay, K., Simon, N., and Qian, J. (2015). Lasso and Elastic-Net Regularized Generalized Linear Models. https://cran.r-project.org/web/packages/glmnet/glmnet.pdf.

Gao, N., White, P., and Kaestner, K.H. (2009a). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev. Cell *16*, 588–599.

Gao, Z., Ure, K., Ables, J.L., Lagace, D.C., Nave, K.A., Goebbels, S., Eisch, A.J., and Hsieh, J. (2009b). Neurod1 is essential for the survival and maturation of adult-born neurons. Nat. Neurosci. *12*, 1090–1092.

Gao, D., Vela, I., Sboner, A., Iaquinta, P.J., Karthaus, W.R., Gopalan, A., Dowling, C., Wanjala, J.N., Undvall, E.A., Arora, V.K., et al. (2014). Organoid cultures derived from patients with advanced prostate cancer. Cell *159*, 176–187.

Gao, S., Yan, L., Wang, R., Li, J., Yong, J., Zhou, X., Wei, Y., Wu, X., Wang, X., Fan, X., et al. (2018). Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. Nat. Cell Biol. *20*, 721–734.

Garrido-Martín, E.M., Blanco, F.J., Roquè, M., Novensà, L., Tarocchi, M., Lang, U.E., Suzuki, T., Friedman, S.L., Botella, L.M., and Bernabéu, C. (2013). Vascular injury triggers Krüppel-like factor 6 mobilization and cooperation with specificity protein 1 to promote endothelial activation through upregulation of the activin receptor-like kinase 1 gene. Circ. Res. *112*, 113–127.

Goding, C.R. (2000). Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage. Genes Dev. *14*, 1712–1728.

Green, M.D., Chen, A., Nostro, M.C., d'Souza, S.L., Schaniel, C., Lemischka, I.R., Gouon-Evans, V., Keller, G., and Snoeck, H.W. (2011). Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. Nat. Biotechnol. *29*, 267–272.

Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G.C., Yang, X., Kirita, Y., and Morris, S.A. (2018). CellTag Indexing: a genetic barcode-based multiplexing tool for single-cell technologies. Genome Biol. *20*, 90.

Han, H., Cho, J.W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018a). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. *46* (D1), D380–D386.

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018b). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell *172*, 1091–1107.e17.

Hirosawa, M., Fujita, Y., Parr, C.J.C., Hayashi, K., Kashida, S., Hotta, A., Woltjen, K., and Saito, H. (2017). Cell-type-specific genome editing with a micro-RNA-responsive CRISPR-Cas9 switch. Nucleic Acids Res. *45*, e118.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features between human and mouse cortex. Nature *573*, 61–68.

Houle, M.E., Kriegel, H.P., Kröger, P., Schubert, E., and Zimek, A. (2010). Can Shared-Neighbor Distances Defeat the Curse of Dimensionality? In Scientific

CellPress

Cell Resource

and Statistical Database Management, M. Gertz and B. Ludäscher, eds. (Springer), pp. 482–500.

Huch, M., and Koo, B.-K. (2015). Modeling mouse and human development using organoid cultures. Development *142*, 3113–3125.

Huch, M., Knoblich, J.A., Lutolf, M.P., and Martinez-Arias, A. (2017). The hope and the hype of organoid research. Development *144*, 938–941.

Jabaudon, D., and Lancaster, M. (2018). Exploring landscapes of brain morphogenesis with organoids. Development *145*, dev172049.

Jang, S., Choubey, S., Furchtgott, L., Zou, L.N., Doyle, A., Menon, V., Loew, E.B., Krostag, A.R., Martinez, R.A., Madisen, L., et al. (2017). Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. eLife 6, 1–28.

Jung, P., Sato, T., Merlos-Suárez, A., Barriga, F.M., Iglesias, M., Rossell, D., Auer, H., Gallardo, M., Blasco, M.A., Sancho, E., et al. (2011). Isolation and in vitro expansion of human colonic stem cells. Nat. Med. *17*, 1225–1227.

Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. J. Clin. Invest. *119*, 1420–1428.

Khalaf-Nazzal, R., Stouffer, M.A., Olaso, R., Muresan, L., Roumegous, A., Lavilla, V., Carpentier, W., Moutkine, I., Dumont, S., Albaud, B., et al. (2017). Early born neurons are abnormally positioned in the doublecortin knockout hippocampus. Hum. Mol. Genet. *26*, 90–108.

Kim, T., and Shivdasani, R.A. (2016). Stomach development, stem cells and disease. Development *143*, 554–565.

Kim, J.W., Botvinnik, O.B., Abudayyeh, O., Birger, C., Rosenbluh, J., Shrestha, Y., Abazeed, M.E., Hammerman, P.S., DiCara, D., Konieczkowski, D.J., et al. (2016). Characterizing genomic alterations in cancer by complementary functional associations. Nat. Biotechnol. *34*, 539–546.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for singlecell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of Tissue-Specific MicroRNAs from Mouse. Curr. Biol. *12*, 735–739.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

Lensch, M.W., Schlaeger, T.M., Zon, L.I., and Daley, G.Q. (2007). Teratoma formation assays with human embryonic stem cells: a rationale for one type of human-animal chimera. Cell Stem Cell *1*, 253–258.

Li, W., and Pozzo-Miller, L. (2014). BDNF deregulation in Rett syndrome. Neuropharmacology 76, 737–746.

Li, J., Huang, H., Sun, L., Yang, M., Pan, C., Chen, W., Wu, D., Lin, Z., Zeng, C., Yao, Y., et al. (2009). MiR-21 indicates poor prognosis in tongue squamous cell carcinomas as an apoptosis inhibitor. Clin. Cancer Res. *15*, 3998–4008.

Lin, Y., Chen, D., Fan, Q., and Zhang, H. (2009). Characterization of SoxB2 and SoxC genes in amphioxus (Branchiostoma belcheri): implications for their evolutionary conservation. Sci. China C Life Sci. *52*, 813–822.

Liu, C., Oikonomopoulos, A., Sayed, N., and Wu, J.C. (2018). Modeling human diseases with induced pluripotent stem cells: from 2D to 3D and beyond. Development *145*, dev156166.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., and Ferrando, A.A. (2005). MicroRNA expression profiles classify human cancers. Nature *435*, 834–838.

Lu, Z., Liu, M., Stribinskis, V., Klinge, C.M., Ramos, K.S., Colburn, N.H., and Li, Y. (2008). MicroRNA-21 promotes cell transformation by targeting the programmed cell death 4 gene. Oncogene *27*, 4373–4379.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202–1214.

Marmigère, F., Montelius, A., Wegner, M., Groner, Y., Reichardt, L.F., and Ernfors, P. (2006). The Runx1/AML1 transcription factor selectively regulates development and survival of TrkA nociceptive sensory neurons. Nat. Neurosci. 9, 180–187.

McInnes, L., and Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv, arXiv:1802.03426. https://arxiv.org/abs/1802.03426.

Medina, P.P., and Slack, F.J. (2008). MicroRNAs and cancer: An overview. Cell Cycle 7, 2485–2492.

Miki, K., Endo, K., Takahashi, S., Funakoshi, S., Takei, I., Katayama, S., Toyoda, T., Kotaka, M., Takaki, T., Umeda, M., et al. (2015). Efficient detection and purification of cell populations using synthetic micro-RNA switches. Cell Stem Cell *16*, 699–711.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature 582, 410–415.

Mort, R.L., Jackson, I.J., and Patton, E.E. (2015). The melanocyte lineage in development and disease. Development *142*, 620–632.

Nissim, L., Wu, M.R., Pery, E., Binder-Nissim, A., Suzuki, H.I., Stupp, D., Wehrspaun, C., Tabach, Y., Sharp, P.A., and Lu, T.K. (2017). Synthetic RNA-based immunomodulatory gene circuits for cancer immunotherapy. Cell *171*, 1138– 1150.e15.

Ortmann, D., and Vallier, L. (2017). Variability of human pluripotent stem cell lines. Curr. Opin. Genet. Dev. 46, 179–185.

Parekh, U., Wu, Y., Zhao, D., Worlikar, A., Shah, N., Zhang, K., and Mali, P. (2018). Mapping Cellular Reprogramming via Pooled Overexpression Screens with Paired Fitness and Single-Cell RNA-Sequencing Readout. Cell Syst. 7, 548–555.e8.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970–985.

Philipp, F., Selich, A., Rothe, M., Hoffmann, D., Rittinghausen, S., Morgan, M.A., Klatt, D., Glage, S., Lienenklaus, S., Neuhaus, V., et al. (2018). Human Teratoma-Derived Hematopoiesis Is a Highly Polyclonal Process Supported by Human Umbilical Vein Endothelial Cells. Stem Cell Reports *11*, 1051–1060.

Phipson, B., Er, P.X., Combes, A.N., Forbes, T.A., Howden, S.E., Zappia, L., Yen, H.J., Lawlor, K.T., Hale, L.J., Sun, J., et al. (2019). Evaluation of variability in human kidney organoids. Nat. Methods *16*, 79–87.

Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 566, 490–495.

Pinello, L., Canver, M.C., Hoban, M.D., Orkin, S.H., Kohn, D.B., Bauer, D.E., and Yuan, G.C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. Nat. Biotechnol. *34*, 695–697.

Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. Neuron.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Qin, Q., Xu, Y., He, T., Qin, C., and Xu, J. (2012). Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. Cell Res. *22*, 90–106.

Quadrato, G., Nguyen, T., Macosko, E.Z., Sherwood, J.L., Min Yang, S., Berger, D.R., Maria, N., Scholvin, J., Goldman, M., Kinney, J.P., et al. (2017).





Cell diversity and network dynamics in photosensitive human brain organoids. Nature *545*, 48–53.

Que, J., Okubo, T., Goldenring, J.R., Nam, K.T., Kurotani, R., Morrisey, E.E., Taranova, O., Pevny, L.H., and Hogan, B.L. (2007). Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. Development *134*, 2521–2531.

Raff, R.A. (1996). The Shape of Life: Genes, Development and the Evolution of Animal Form (University of Chicago Press).

Richardson, M.K., Hanken, J., Gooneratne, M.L., Pieau, C., Raynaud, A., Selwood, L., and Wright, G.M. (1997). There is no highly conserved embryonic stage in the vertebrates: implications for current theories of evolution and development. Anat. Embryol. *196*, 91–106.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Mukherjee, S., Chen, W., Peeler, D.J., Yao, Z., Tasic, B., Sellers, D.L., et al. (2017). Scaling single cell transcriptomics through split pool barcoding. bioRxiv. https://doi.org/10. 1101/105163.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.

Royo, J.L., Maeso, I., Irimia, M., Gao, F., Peter, I.S., Lopes, C.S., D'Aniello, S., Casares, F., Davidson, E.H., Garcia-Fernández, J., and Gómez-Skarmeta, J.L. (2011). Transphyletic conservation of developmental regulatory state in animal evolution. Proc. Natl. Acad. Sci. USA *108*, 14186–14191.

Samatov, T.R., Wicklein, D., and Tonevitsky, A.G. (2016). L1CAM: Cell adhesion and more. Prog. Histochem. Cytochem. *51*, 25–32.

Sammon, J.W. (1969). A Nonlinear Mapping for Data Structure Analysis. IEEE Trans. Comput. *C*–18, 401–409.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Sarper, S.E., Inubushi, T., Kurosaka, H., Ono Minagi, H., Kuremoto, K.I., Sakai, T., Taniuchi, I., and Yamashiro, T. (2018). Runx1-Stat3 signaling regulates the epithelial stem cells in continuously growing incisors. Sci. Rep. 8, 10906.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Sato, T., Stange, D.E., Ferrante, M., Vries, R.G., Van Es, J.H., Van den Brink, S., Van Houdt, W.J., Pronk, A., Van Gorp, J., Siersema, P.D., and Clevers, H. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology *141*, 1762–1772.

Scheitz, C.J.F., and Tumbar, T. (2013). New insights into the role of Runx1 in epithelial stem cell biology and pathology. J. Cell. Biochem. *114*, 985–993.

Shapiro, B., Tocci, P., Haase, G., Gavert, N., and Ben-Ze'ev, A. (2015). Clusterin, a gene enriched in intestinal stem cells, is required for L1-mediated colon cancer metastasis. Oncotarget *6*, 34389–34401.

Shi, M., Kovac, A., Korff, A., Cook, T.J., Ginghina, C., Bullock, K.M., Yang, L., Stewart, T., Zheng, D., Aro, P., et al. (2016). CNS tau efflux via exosomes is likely increased in Parkinson's disease but not in Alzheimer's disease. Alzheimers Dement. *12*, 1125–1131.

Shivdasani, R.A. (2006). MicroRNAs: regulators of gene expression and cell differentiation. Blood *108*, 3646–3654.

Silberg, D.G., Sullivan, J., Kang, E., Swain, G.P., Moffett, J., Sund, N.J., Sackett, S.D., and Kaestner, K.H. (2002). Cdx2 Ectopic Expression Induces Gastric Intestinal Metaplasia in Transgenic Mice. Gastroenterology *122*, 689–696.

Simmini, S., Bialecka, M., Huch, M., Kester, L., van de Wetering, M., Sato, T., Beck, F., van Oudenaarden, A., Clevers, H., and Deschamps, J. (2014). Transformation of intestinal stem cells into gastric stem cells on loss of transcription factor Cdx2. Nat. Commun. *5*, 5728.

Smirnova, L., Gräfe, A., Seiler, A., Schumacher, S., Nitsch, R., and Wulczyn, F.G. (2005). Regulation of miRNA expression during neural cell specification. Eur. J. Neurosci. *21*, 1469–1477.

Smith, K.P., Luong, M.X., and Stein, G.S. (2009). Pluripotency: toward a gold standard for human ES and iPS cells. J. Cell. Physiol. *220*, 21–29.

Stevens, L.C. (1962). The biology of teratomas including evidence indicating their origin form primordial germ cells. Annee Biol. 1, 585–610.

Stevens, L.C. (1967). The biology of teratomas. Adv. Morphog. 6, 1-31.

Stevens, L.C., and Pierce, G.B. (1975). Teratomas: Definitions and Terminology. In Teratomas and Differentiation, M.I. Sherman and D. Solter, eds. (Academic Press), pp. 13–14.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Stumpel, C., and Vos, Y.J. (1993). L1 Syndrome. In GeneReviews, M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K. Stephens, and A. Amemiya, eds. (University of Washington).

Sun, Y., Luo, Z.M., Guo, X.M., Su, D.F., and Liu, X. (2015). An updated role of microRNA-124 in central nervous system disorders: a review. Front. Cell. Neurosci. *9*, 193.

Suzuki, N., Yamazaki, S., Yamaguchi, T., Okabe, M., Masaki, H., Takaki, S., Otsu, M., and Nakauchi, H. (2013). Generation of engraftable hematopoietic stem cells from induced pluripotent stem cells by way of teratoma formation. Mol. Ther. *21*, 1424–1431.

Suzuki, A., Shinoda, M., Honda, K., Shirakawa, T., and Iwata, K. (2016). Regulation of transient receptor potential vanilloid 1 expression in trigeminal ganglion neurons via methyl-CpG binding protein 2 signaling contributes tongue heat sensitivity and inflammatory hyperalgesia in mice. Mol. Pain *12*, 1–11.

Tarlow, D., Swersky, K., Charlin, L., Sutskever, I., and Zemel, R. (2013). Stochastic k-Neighborhood Selection for Supervised and Unsupervised Learning. Proceedings of Machine Learning Research *28*, 199–207.

Thissen, D., Steinberg, L., and Kuang, D. (2002). Quick and Easy Implementation of the Benjamini-Hochberg Procedure for Controlling the False Positive Rate in Multiple Comparisons. Journal of Educational and Behavioral Statistics *27*, 77–83.

Thurlbeck, W.M., and Scully, R.E. (1960). Solid Teratoma of the Ovary. A Clinicopathological Analysis of 9 Cases. Cancer *13*, 804–811.

Tsukada, M., Ota, Y., Wilkinson, A.C., Becker, H.J., Osato, M., Nakauchi, H., and Yamazaki, S. (2017). In Vivo Generation of Engraftable Murine Hematopoietic Stem Cells by Gfi1b, c-Fos, and Gata2 Overexpression within Teratoma. Stem Cell Reports *9*, 1024–1033.

Tsunemoto, R., Lee, S., Szűcs, A., Chubukov, P., Sokolova, I., Blanchard, J.W., Eade, K.T., Bruggemann, J., Wu, C., Torkamani, A., et al. (2018). Diverse reprogramming codes for neuronal identity. Nature *557*, 375–380.

Umansky, K.B., Gruenbaum-Cohen, Y., Tsoory, M., Feldmesser, E., Goldenberg, D., Brenner, O., and Groner, Y. (2015). Runx1 Transcription Factor Is Required for Myoblasts Proliferation during Muscle Regeneration. PLoS Genet. *11*, e1005457.

van de Wetering, M., Francies, H.E., Francis, J.M., Bounova, G., Iorio, F., Pronk, A., van Houdt, W., van Gorp, J., Taylor-Weiner, A., Kester, L., et al. (2015). Prospective derivation of a living organoid biobank of colorectal cancer patients. Cell *161*, 933–945.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27.

Vastag, L., Jorgensen, P., Peshkin, L., Wei, R., Rabinowitz, J.D., and Kirschner, M.W. (2011). Remodeling of the metabolome during early frog development. PLoS ONE 6, e16881.

Velasco, S., Kedaigle, A.J., Simmons, S.K., Nash, A., Rocha, M., Quadrato, G., Paulsen, B., Nguyen, L., Adiconis, X., Regev, A., et al. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. Nature 570, 523–527.

Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science *360*, 981–987.

CelPress



Wang, S., Aurora, A.B., Johnson, B.A., Qi, X., McAnally, J., Hill, J.A., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2008). The Endothelial-Specific MicroRNA miR-126 Governs Vascular Integrity and Angiogenesis. Dev. Cell *15*, 261–271.

Wang, J., Jenjaroenpun, P., Bhinge, A., Angarica, V.E., Del Sol, A., Nookaew, I., Kuznetsov, V.A., and Stanton, L.W. (2017). Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons. Genome Res. *27*, 1783–1794.

Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. J. Econ. Entomol. *39*, 269–270.

Willis, R.A. (1934). The Structure of Teratoma. J. Pathol. Bacteriol. XL.

Willis, R.A. (1936). THE HISTOGENESIS OF NEURAL TISSUE IN TERA-TOMAS. J. Pathol. Bacteriol. 42, 411–416.

Wolburg, H., Wolburg-Buchholz, K., Mack, A.F., and Reichenbach, A. (2009). Ependymal cells. Encyclopedia of Neuroscience, 1133–1140.

Wu, Y., Tamayo, P., and Zhang, K. (2018). Visualizing and Interpreting Single-Cell Gene Expression Datasets with Similarity Weighted Nonnegative Embedding. Cell Syst. 7, 656–666.e4.

Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927–939.

Yao, Q., Xu, H., Zhang, Q.Q., Zhou, H., and Qu, L.H. (2009). MicroRNA-21 promotes cell proliferation and down-regulates the expression of programmed cell death 4 (PDCD4) in HeLa cervical carcinoma cells. Biochem. Biophys. Res. Commun. 388, 539–542.

Yao, Z., Mich, J.K., Ku, S., Menon, V., Krostag, A.R., Martinez, R.A., Furchtgott, L., Mulholland, H., Bort, S., Fuqua, M.A., et al. (2017). A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development. Cell Stem Cell *20*, 120–134.

Yin, X., Mead, B.E., Safaee, H., Langer, R., Karp, J.M., and Levy, O. (2016). Engineering Stem Cell Organoids. Cell Stem Cell *18*, 25–38.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049.

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature 555, 524–528.

Zhu, S., Wu, H., Wu, F., Nie, D., Sheng, S., and Mo, Y.Y. (2008). MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. Cell Res. *18*, 350–359.

Zhu, Y., Sousa, A.M.M., Gao, T., Skarica, M., Li, M., Santpere, G., Esteller-Cucala, P., Juan, D., Ferrández-Peral, L., Gulden, F.O., et al. (2018). Spatiotemporal transcriptomic divergence across human and macaque brain development. Science *362*, eaat8077.

Zovein, A.C., Hofmann, J.J., Lynch, M., French, W.J., Turlo, K.A., Yang, Y., Becker, M.S., Zanetta, L., Dejana, E., Gasson, J.C., et al. (2008). Fate tracing reveals the endothelial origin of hematopoietic stem cells. Cell Stem Cell *3*, 625–636.

Cell Resource



STAR***METHODS**

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER	
Antibodies			
Anti-Pax6 (rabbit polyclonal)	Millipore Sigma	AB2237 Lot: 3324991 RRID:AB_1587367	
Dnk pAb to Rb IgG Dylight® 550	Abcam	ab98489 Lot: GR33528-3 RRID:AB_10674190	
anti-rabbit IgG Alexa 488	Invitrogen	Ref: A11008 Lot: 2179202 RRID:AB_143165	
Bacterial and Virus Strains			
One Shot® Stbl3	Invitrogen	Ref: C7373-03	
Biological Samples			
N/A	N/A	N/A	
Chemicals, Peptides, and Recombinant Proteins			
Lipofectamine 2000	Invitrogen	Ref: 11668-019 Lot: 2196156	
Critical Commercial Assays			
Chromium Single Cell 3' Reagent Kit	10x Genomics	PN-1000075	
Tumor Dissociation Kit, Human	Miltenyi Biotec	Order No. 130-095-929	
RNAScope® Fluorescent Multiplex Reagent Kit V1	Advanced Cell Diagnostics	Cat# 320850	
Deposited Data			
Raw and processed data	This paper	GEO: GSE156170	
Mouse Cell Atlas	Han et al., 2018b	GEO: GSE108097	
Mouse Organogenesis Cell Atlas	Cao et al., 2019	GEO: GSE119945	
Fetal Human Cortex scRNA-seq Dataset (gw 5 -gw18)	Zhong et al., 2018	GEO: GSE104276	
Fetal Human Cortex scRNA-seq Dataset (gw 17 -gw18)	Polioudakis et al., 2019	http://solo.bmap.ucla.edu/shiny/webapp/, dbGAP: phs001836	
Fetal Human Gut scRNA-seq Dataset	Gao et al., 2018	GEO: GSE103239	
Combined Human (hg19) and Mouse (mm10) genome reference and gene annotations	10X Genomics	https://support.10xgenomics.com/ single-cell-gene-expression/software/ pipelines/latest/advanced/references	
Gene Ontology Biological Process Genesets	MSigDB	https://www.gsea-msigdb.org/gsea/ msigdb/collections.jsp#C5	
TRRUST TF database	Han et al., 2018a	https://www.grnpedia.org/trrust/	
Experimental Models: Cell Lines			
H1 Embryonic Stem Cell Line	WiCell	RRID:CVCL_N541	
H9 Embryonic Stem Cell Line	WiCell	RRID:CVCL_1240	
HUES62 Embryonic Stem Cell Line	WiCell	RRID:CVCL_B197	
PGP1 Induced Pluripotent Stem Cell Line	Dr. George Church	RRID:CVCL_F182	
HEK293T	ATCC	RRID:CVCL_0063	
HUVEC	ATCC	RRID:CVCL_2959	
HeLa	ATCC	RRID:CVCL_0030	
Experimental Models: Organisms/Strains			
Male NOD-scid IL2Rgamma ^{null} Mice	UC San Diego ACP	RRID:BCBC_4142	
Oligonucleotides			
(See Table S7 and Star Methods)	This paper	N/A	

(Continued on next page)

CellPress



Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
CROPseq-Guide-Puro	Datlinger et al., 2017	CROPseq-Guide-Puro Cat# 86708
Lenti-guide-puro	Sanjana et al., 2014	Lenti-guide-puro Cat# 52963
pCR4-Blunt-TOPO	Thermo Fisher Scientific	Cat# K2875J10
CAG-spCas9-P2A-EGFP	This paper	N/A
Ef1alpha-mCherry-IRES-Hygro-Barcode	This paper	N/A
miRNA-HSV-tk-GFP	This paper	N/A
pMD2.G	Trono Lab	pMD2.G Cat# 12259
pCMV delta R8.2	Trono Lab	pCMV delta R8.2 Cat# 12263
Software and Algorithms		
Code for processing and analyzing teratoma data	This paper	https://www.github.com/yanwu2014/ teratoma-analysis-code
R/RStudio	R/RStudio	https://rstudio.com/
CellRanger	10X Genomics	https://support.10xgenomics.com/ single-cell-gene-expression/software/ pipelines/latest/what-is-cell-ranger
Seurat	Stuart et al., 2019	https://satijalab.org/seurat/
SWNE	Wu et al., 2018	https://github.com/yanwu2014/swne
perturbLM	Parekh et al., 2018	https://github.com/yanwu2014/perturbLM
phEMD	Chen et al., 2020	https://www.bioconductor.org/packages/ release/bioc/html/phemd.html
CellMapper	Github	https://github.com/yanwu2014/cellMapper
ImageJ (FIJI)	ImageJ Team	https://imagej.net/Welcome
Genotyping Matrices	Parekh et al., 2018	https://github.com/yanwu2014/ genotyping-matrices
PicardTools	Broad Institute	https://broadinstitute.github.io/picard/
Other		
Ganciclovir (99.77%)	MedChemExpress	Cat# HY-13637

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Prashant Mali (pmali@ucsd.edu).

Materials Availability

Until the Addgene submission process completes, all unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

Data and Code Availability

The accession number for the scRNA-seq data reported in this paper is GEO: GSE156170

All code used for analysis are available at this github repository: yanwu2014/teratoma-analysis-code. Instructions for reproducing our analysis step by step are also in this repository.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture

The H1 (P30), H9 (P36), PGP1 (P39), and HUES62 (P20) hESC cell line was maintained under feeder-free conditions in mTeSR medium (Stem Cell Technologies). Prior to passaging, tissue-culture plates were coated with growth factor-reduced Matrigel (Corning) diluted in DMEM/F-12/Glutamax medium (Thermo Fisher Scientific), and incubated for 30 minutes at 37°C, 5% CO2. Cells were





dissociated and passaged using the dissociation reagent Versene (Thermo Fisher Scientific). Cells were passaged a maximum of 4 times for proper expansion prior to injection. HEK293T and HeLa were maintained in high glucose DMEM supplemented with 10% fetal bovine serum (FBS) and passaged every couple days upon confluency with 0.05% Trypsin-EDTA (GIBCO). HUVECs were maintained in EGM-2 (Lonza).

Organoid Generation and Dissociation

Self-patterned whole brain organoids were generated following the Quadrato et al., 2017 protocol (Quadrato et al., 2017). Briefly, H1 ESCs transduced with either miR-124-HSV-tk-GFP or HSV-tk-GFP were cultured as embryoid bodies for 5 days, transferred into Neural Induction (NI) media for 5 days, and finally embedded in Matrigel and cultured in Cortical Differentiation (CD) media for 25 days. Day 35 organoids were dissociated to single cell following a modified protocol using the GentleMACS Human Tumor Dissociation Kit, but without use of the GentleMACS dissociator and instead cells were triturated post- 37° C 1-hr incubation with a 1000 µL pipetman prior to 70 µM filtration. Resulting single cell suspension was analyzed for GFP florescence via flow cytometry. Cells, embryoid bodies, and organoids were maintained under puromycin selection [0.75µg/µL] for the entirety of the experiment.

Animals

Animals used in this study were male NOD-scid IL2Rgamma^{null} mice 8-10 weeks of age. Housing, husbandry and all procedures involving animals used in this study were performed in compliance with protocols (#S16003) approved by the University of California San Diego Institutional Animal Care and Use Committee (UCSD IACUC). Mice were group housed (up to 4 animals per cage) on a 12:12 hr light-dark cycle, with free access tso food and water in individually ventilated specific pathogen free (SPF) autoclaved cages. All mice used were healthy and were not involved in any previous procedures nor drug treatment unless indicated otherwise.

METHOD DETAILS

PGP1-Cas9 Clone Generation

The PGP1 human induced pluripotent stem cell line was a kind gift of Dr. George Church at Harvard Medical School. The sgRNA targeting AAVS1 locus of the human genome (spacer sequence GGGCCACTAGGGACAGGAT) was cloned into the Lenti-guidepuro plasmid (Addgene #52963). To generate the knockin donor plasmid, we cloned the CAG promoter followed by a cassette of co-expression of spCas9 and EGFP splitting via the P2A sequence into the pCR4-Blunt-TOPO vector (Thermo Fisher Scientific). Two homology arms were amplified from upstream (804 bp) and downstream (837 bp) of the sgRNA targeting site in AAVS1 genomic locus and constructed into the donor plasmid flanking the CAG-spCas9-P2A-EGFP cassette. Between the upstream homology arm and the CAG promoter, we inserted a splice acceptor sequence following by a T2A linked blasticidin resistance gene.

Human iPSC PGP1 cells were electroporated using 4D-Nucleofector system and P3 Primary Cell X kit (Lonza) according to the manufacturer's instruction. Briefly, the PGP1 cells were dissociated into single cells. 1×10^6 cells were mixed with 100 µL nucleofection reagents and 10 µg DNA (5 µg Cas9 donor + 5 µg sgRNA) and electroporated. The cells were recovered with pre-warmed medium and then cultured on inactivated MEF feeders in 10 cm dishes with mTeSR medium supplemented with 0.5 µM ROCK-inhibitor. Afterward, the mTeSR medium without ROCK-inhibitor was refreshed daily. 2 µg/ml blasticidin were added into the culture medium 7 days after electroporation. The cells were cultured without passage until clones emerged on the plate. The clones were checked under the microscope and those with EGFP expression were picked up and expanded individually.

To detect genomic integration, the genomic DNA from cultured cells was extracted using DNeasy Blood & Tissue Kits (QIAGEN). Approximately 500 ng of genomic DNA was used for each PCR reaction using KAPA HiFi HotStart Ready Mix (Kapa Biosystems). The PCR amplification of the left and right arm utilized primers that amplified regions spanning both the PGP1 AAVS1 endogenous locus and the engineered cassette (Figure S4B).

The primer sequences are listed below.

Left_arm_forward	ACTTCCCCTCTTCCGATGTTG
Left_arm_reverse	ATTGTAGCCGTTGCTCTTTCA
Right_arm_forward	GAGCAAAGACCCCAACGAGAAGC
Right_arm_reverse	CTGCCTGGAGAAGGATGCAGGA

This was further validated by direct Sanger sequencing of the arms (Figure S4A), The activity of Cas9 in the PGP1-Cas9 cells was validated by the generation of indels at the expected position when guide RNAs were introduced.

sgRNA Design

The CRISPR-KO sgRNA sequences targeting transcription factor genes were obtained from the GPP sgRNA Designer web tool (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design, accessed February 2018) as follows. The 24 gene symbols in the table below were converted to Entrez gene IDs using Bioconductor package org.Hs.eg.db_3.5.0, and the resulting IDs

CellPress



were submitted together with the following parameters: enzyme Sp, taxon human, quota 50, include unpicked. From the resulting output, the two guide sequences with the highest "pick order" were selected for each target gene. To check the validity of each guide sequence, the corresponding context sequence was compared to the human reference genome at the predicted cut location using Bioconductor package BSgenome.Hsapiens.UCSC.hg38_1.4.1, and the cut location was confirmed to be fully within the target gene coding sequence determined using Bioconductor package TxDb.Hsapiens.UCSC.hg38.knownGene_3.4.0.

Gene symbol		Entre	z ID	sgRNA-1		sgRNA-2
SOX17		64321	l	GGCAACGGG	TAGCCGTCGAG	AGGGCGAGTCCCGTATCCGG
CDX2		1045		CCGCAGTACC	CGGACTACGG	CAAATATCGAGTGGTGTACA
HNF4A		3172		GGGACCGGAT	TCAGCACTCGA	GCAATGACTACATTGTCCCT
GATA4		2626		TGTGGGCACG	TAGACTGGCG	CCGGCTTACATGGCCGACGT
GATA6		2627		CGGGACGCCT	TCAGCTCGACA	GCCGACAGCGAGCTGTACTG
RUNX1		861		CTGATCGTAG	GACCACGGTG	TGCTCCCCACAATAGGACAT
FOXA2		3170		ATGAACATGT	CGTCGTACGT	TCCGTGAGCAACATGAACGC
PDX1		3651		GGAGAACAAG	GCGGACGCGCA	TATTCAACAAGTACATCTCA
NKX2-1		7080		GCGAGCGGC	ATGAACATGAG	GGTTGGCGCCGTACCATCCG
NKX2-5		1482		GTAGGCACGT	GGATAGAAGG	GAAGACAGAGGCGGACAACG
SOX9		6662		ACGTCGCGGA	AGTCGATAGG	TTCACCGACTTCCTCCGCCG
PROX1		5629		AGTGTCCACA	ACTTGCGACA	CGGGTTGAGAATATAATTCG
SNAI1		6615		GGGACTCTCC	TGGAGCCGAA	TGTAGTTAGGCTTCCGATTG
TWIST1		7291		CGGGAGTCCC	GCAGTCTTACG	AGCGGGTCATGGCCAACGTG
ASCL1		429		CCAGGTTGAC	CAACTTGACG	AAACGCCGGCTCAACTTCAG
NEUROG1		4762		CCGCATGCAC	AACTTGAACG	TTGGTGTCGTCGGGGGAACGA
KLF6		1316		TCTGAGGCTG	AAACATAGCA	GCTGACCAAAACTTCGCCAA
KLF2		10365	5	GGTTCGGGGT	TAATAGAACGC	CTTCGGTCTCTTCGACGACG
HES1		3280		GTGCGAGGG	CGTTAATACCG	AGCCAGTGTCAACACGACAC
FOXG1		2290		AGCGCGTTGT	AGCTGAACGG	CCGCGCCACTACGACGACCC
TULP3		7289		GGAGTATGAC	AGTTCACCAA	TGAAAGTGTGAACTTCGATG
MYOG		4656		TTACACACCT	FACACGCCCA	TCGAACCACCAGGCTACGAG
GATA3		2625		TCCAAGACGT	CCATCCACCA	CAGGGAGTGTGTGAACTGTG
FGFR2		2263		CTTAGTCCAA	CTGATCACGG	TGACCAAACGTATCCCCCTG
Gene symbol	Entrez ID		sgRNA-1		sgRNA-2	sgRNA-3
TCF4	6925		GTGGACATCGGAG	GAAGAC	TGTCCACTTTCCATCGTAG	CAAACGTTCATGTGGATGC
MECP2	4204		GCTCCATCATCCGT	GACCG	AAAGCCTTTCGCTCTAAAG	TTGCGTACTTCGAAAAGGT
L1CAM	3897		GCGTCCGGTGTCAT	TTGGCC	GCGTACTATGTCACCGTGG	GCCAGTACCGAACTGGATG

Library Preparation

The lentiviral backbone plasmid for the barcode vector was constructed containing the EF1 α promoter, mCherry transgene flanked by BamHI restriction sites, followed by a P2A peptide and hygromycin resistance enzyme gene immediately downstream (ECIH). The backbone was digested with HpaI, and a pool of 20 bp long barcodes with flanking sequences compatible with the HpaI site, was inserted immediately downstream of the hygromycin resistance gene by Gibson assembly. The vector was constructed such that the barcodes were located only 200 bp upstream of the 3'-LTR region. This design enabled the barcodes to be transcribed near the polyadenylation tail of the transcripts and a high fraction of barcodes to be captured during sample processing for scRNA-seq.

The lentiviral backbone plasmid for the sgRNAs was the CROPseq-Guide-Puro vector (Addgene #86708). To create the sgRNA library, individual sgRNAs were PCR amplified utilizing overlapping forward and reverse primers custom designed with flanking sequences compatible with the BSMBI restriction sites (Table S4B). The lentiviral backbone was digested with BSMBI (New England Biolabs) at 55°C for 3 hours in a reaction consisting of: CROPseq-Guide-Puro backbone, 5 µg, Buffer NEB 3.1, 5 µl, BSMBI, 5 µl, H20 up to 50 µl. After digestion, the vector was purified using a QIAquick PCR Purification Kit (QIAGEN). Each sgRNA was then individually assembled via Gibson assembly.

The lentiviral backbone plasmid for the miRNA-HSV-tk-GFP constructs was an EF1-alpha promoter, GFP, IRES domain, and puromycin-resistance gene (EGIP) backbone. The lentiviral backbone was digested with EcoRV-HF (New England Biolabs) at 37°C for 1





hour to excise out the GFP in a reaction consisting of: EGIP backbone, 5 μg, 1X Cutsmart Buffer (New England Biolabs), 5 μl, EcoRV-HF, 5 μl, H20 up to 50 μl. After digestion, the vector was purified using a QIAquick PCR Purification Kit (QIAGEN). We amplified a gBlock containing the Herpes Simplex Virus thymidine kinase (HSV-tk), 2A self-cleaving peptide, and GFP. The primers used to amplify the gBlock contain unique miRNA binding sites (see below).

miR_Empty_F	TGGCTAGTTAAGCTTGATATCGA ATTCCTGCAGCCCGGGGGATCCA GATCACACCGGTCGCCA
miR_Empty_R	GGGAGAGGGGGGGGGGGGGGG AATTCCGCGGGCCCGTCGAC GCGGTTAACGCCGCTTTACTTGTACAG
miR_21_F	TGGCTAGTTAAGCTTGATATCGAATT CCTGCAGCCCGGGGGATCCTCAACA TCAGTCTGATAAGCTA AGATC ACACCGGTCGCCA
miR_21_R	GGGAGAGGGGGGGGGGGGGGGGAATT CCGCGGGCCCGTCGACGCGGTTTAG CTTATCAGACTGATGTTGA AACG CCGCTTTACTTGTACAG
miR_122_F	TGGCTAGTTAAGCTTGATATCGAATTC CTGCAGCCCGGGGGGATCCCAAACAC CATTGTCACACTCCA AGAT CACACCGGTCGCCA
miR_122_R	GGGAGAGGGGGGGGGGGGGGGAA TTCCGCGGGCCCGTCGACGCGG TTTGGAGTGTGACAATGGTGTTTG AACGCCGCTTTACTTGTACAG
miR_124_F	TGGCTAGTTAAGCTTGATATC GAATTCCTGCAGCCCGGGGGA TCCGGCATTCACCGCGTGC CTTA AGATCACACCGGTCGCCA
miR_124_R	GGGAGAGGGGGGGGGGGGGGGGAATT CCGCGGGCCCGTCGACGCGGTTTA AGGCACGCGGTGAATGCC AAC GCCGCTTTACTTGTACAG
miR_126_F	TGGCTAGTTAAGCTTGATATCGA ATTCCTGCAGCCCGGGGGATCC CGCATTATTACTCACGGTAC GA AGATCACACCGGTCGCCA
miR_126_R	GGGAGAGGGGGGGGGGGGGGGAA TTCCGCGGGCCCGTCGACGCGG TTTCGTACCGTGAGTAATAATGCG AACGCCGCTTTACTTGTACAG
miR_302A_F	TGGCTAGTTAAGCTTGATATCGAA TTCCTGCAGCCCGGGGGATCCAGC AAGTACATCCACGTTTAAGT AGATCACACCGGTCGCCA
miR_302A_R	GGGAGAGGGGGGGGGGGGGGGGAAT TCCGCGGGCCCGTCGACGCGG TTACTTAAACGTGGATGTACTTGC T AACGCCGCTTTACTTGTACAG

We cloned this amplicon into our digested EGIP backbone using standard Gibson assembly.

The Gibson assembly reactions were set up as follows: 1:10 molar ratio of digested backbone to sgRNA insert, 2X Gibson assembly master mix (New England Biolabs), H20 up to 20 µl. After incubation at 50°C for 1 h, the product was transformed into One Shot Stbl3 chemically competent *Escherichia coli* (Invitrogen). A fraction (150 µL) of cultures was spread on carbenicillin (50 µg/ml) LB plates and incubated overnight at 37°C for 15-18hrs (miRNA constructs required longer incubation times). Individual colonies

CellPress



were picked, introduced into 5 mL of carbenicillin (50 μ g/ml) LB medium and incubated overnight in a shaker at 37°C. The plasmid DNA was then extracted with a QIAprep Spin Miniprep Kit (QIAGEN), and Sanger sequenced to verify correct assembly of the vector and to extract barcode sequences.

To assemble the library, individual sgRNA vectors were pooled together in an equal mass ratio along with 5 non-targeting control (NTC) sgRNAs which constituted 50% of the final pool.

Viral Production

HEK293T cells were maintained in high glucose DMEM supplemented with 10% fetal bovine serum (FBS). Cells were seeded in a 15 cm dish 1 day prior to transfection, such that they were 60%-70% confluent at the time of transfection. For each 15 cm dish 36 µL of Lip-ofectamine 2000 (Life Technologies) was added to 1.5 mL of Opti-MEM (Life Technologies). Separately 3 µg of pMD2.G (Addgene #12259), 12 µg of pCMV delta R8.2 (Addgene #12263) and 9 µg of an individual vector or pooled vector library was added to 1.5 mL of Opti-MEM. After 5 minutes of incubation at room temperature, the Lipofectamine 2000 and DNA solutions were mixed and incubated at room temperature for 30 minutes. Medium in each 15 cm dish was replenished with 25 mL of fresh medium. After the incubation period, the mixture was added dropwise to each dish of HEK293T cells. Supernatant containing the viral particles was harvested after 48 and 72 hours, filtered with 0.45 µm filters (Steriflip, Millipore), and further concentrated using Amicon Ultra-15 centrifugal ultrafilters with a 100,000 NMWL cutoff (Millipore) to a final volume of 600-800 µl, divided into aliquots and frozen at -80° C.

Viral Transduction

For viral transduction, virus was added at a low MOI (ensuring a single barcode/cell or a single sgRNA/cell) to stem cells at 20% confluency alongside polybrene (5 μ g/ml, Millipore) in fresh mTeSR medium. The following day, medium was replaced with fresh mTeSR. Appropriate selection reagent was added 48 hr after transduction (hygromycin [50 μ g/ μ L] for barcode / puromycin [0.75 μ g/ μ L] for CRISPR KO screen / miRNA-HSV-tk-GFP) (Thermo Fisher Scientific) and was replaced daily. For miRNA-HSV-tk-GFP transduced cells puromycin selection did not begin until 5-7 days after transduction to allow for enough GFP positive cells. For editing in CRISPR KO screen, selection was continued for 5 days prior to use for teratoma formation in mice.

sgRNA Editing Rate Validation

We individually transduced each sgRNA into our PGP-Cas9 cell line in an arrayed format and selected with puromycin after 48 hr and allowed editing to occur for an additional 5 days (7 days total). From there we retrieved the cell pellets from each individual sgRNA and extracted gDNA. We then designed primers (Table S4C) upstream and downstream of the expected cut site for each individual sgRNA and amplified that region utilizing standard PCR on the gDNA extracted from each cell pellet transduced with each individual sgRNA. Each amplicon for each sgRNA was then sent out for deep sequencing. We used CRISPResso with default parameters to compute the fraction of reads containing mutations, which we split out into an indel rate and an overall mutation rate.

GCV-HSV-tk Killing in vitro

Cells transduced with miRNA-HSV-tk-GFP construct and EGIP-transduced controls grew for a maximum of 5 days in standard medium conditions in the presence of Ganciclovir ([GCV, Sigma-Aldrich] 1 μ M, 10 μ M, or 100 μ M) with daily phase and fluorescent microscopy imaging. GCV was resuspended and stored in 1 mL PBS (GIBCO) aliquots at 3mg/mL in -20° C. Cells were seeded at similar densities on Day 0 of experiment.

miRNA-HSV-tk-GFP Knockdown in vitro

Cells were transduced with miRNA-HSV-tk-GFP constructs and allowed to grow for a maximum of 5 days in standard medium conditions. After 5 days, cells were spun down and resuspended in PBS (GIBCO) at 1x10⁶ cell / mL and ran on the Becton Dickinson FACScan flow cytometer gating for fluorescence (FL1-H [GFP positivity]) and forward scatter (FSC-H [shape and size]).

Teratoma Formation

A subcutaneous injection of 5-10 million PSCs in a slurry of Matrigel® and mTeSR medium (1:1) was made in the right flank of an esthetized Rag2^{-/-}; $\gamma c^{-/-}$ immunodeficient mice. Weekly monitoring of teratoma growth was made by quantifying approximate elliptical area (mm²) with the use of calipers measuring outward width and height.

Molecular Sculpting of Teratomas

Standard teratoma formation protocol was followed using miRNA-HSV-tk-GFP transduced H1s. Once teratomas reach a size of at least 10mm in one axis, intratumoral (IT) or combined intraperitoneal intrautumoral (IPIT) administration of GCV begins at 80mg/kg/ d or 100mg/kg/d (50mg/kg/d at each site) respectively, using standard needle and syringe injection. Teratoma was allowed to grow for a total of 10 weeks before extraction.

Teratoma Processing

After growth for 70 days on average mice were euthanized by slow release of CO2 followed by secondary means via cervical dislocation. Tumor area was shaved, sprayed with 70% ethanol, and then extracted via surgical excision using scissors and forceps.





Tumor was rinsed with PBS, weighed, and photographed. Tumors were inspected for external heterogeneity to ensure proper tumor representation. Representative tumors were cut in a semi-random fashion in ≤ 22 mm diameter pieces and frozen in OCT for sectioning and H&E staining courtesy of the Moore's Cancer Center Histology Core. Remaining tumor was cut into small pieces 1-2mm in diameter and subjected to standard GentlaMACS protocols: Human Tumor Dissociation Kit (medium tumor settings), Red Blood Cell Lysis Kit, and Dead Cell Removal Kit. Single cells were then resuspended in 0.04% BSA for 10X Genomics chromium (Zheng et al., 2017) platform.

Histology and RNAScope®

Sectioning and H&E staining was performed by the Moore's Cancer Center Histology Core. In brief, Optimal Cutting Temperature (O.C.T.) blocks were sectioned with a cryostat into 10 micron sections onto a positively charged glass slide. The slide was then stained with Harris hematoxylin and then rinsed in tap water and treated with an alkaline solution. The slide was then de-stained to remove non-specific background staining with a weak acid alcohol. The section was then stained with an aqueous solution of eosin and passed through several changes of alcohol, then rinsed in several baths of xylene. A thin layer of polystyrene mountant was applied, followed by a glass coverslip. Sections from teratomas were confirmed to have the presence of all 3 germ layers: ectoderm, mesoderm, and endoderm via microscopy identification courtesy of pathologist Dr. Ann Tipps. Further detailed identification also performed by Dr. Tipps.

Fresh frozen sections were subjected to standard RNAScope® Fluorescent Multiplex Reagent Kit protocols following fresh frozen tissue requirements. In brief, sections were fixed with chilled 200 mL of 4% PFA in 1X PBS in 4°C for 15 min. The slides were then placed in 50% EtOH for 5 min at RT, then placed in 70% EtOH for 5 min at RT, and then finally placed in 100% EtOH for 5 min at RT twice. After the slides had dried, we drew a hydrophobic barrier around the tissue. We then placed the dried slides on a HybEZ Slide Rack, and added Pretreat 4 to entirely cover each section and then incubated for 30 min at RT. Slides were then washed with 1X PBS. We then added the appropriate probe to cover each section. Slides were then placed in the slide rack and then placed in a HybEZ Oven for 2 hr at 40°C. After 2 hr, slides were taken out and slides were washed with 1X Wash Buffer for 2 min at RT twice. AMP 1-FL was then added to entirely cover each section. The slides were then placed on the slide rack and inserted into the oven for 30 min at 40°C. The slides were then taken out and slides were washed with 1X Wash Buffer for 2 min at RT twice. AMP 2-FL was then added to entirely cover each section. The slides were then placed on the slide rack and inserted into the oven for 15 min at 40°C. The slides were then taken out and slides were washed with 1X Wash Buffer for 2 min at RT twice. AMP 3-FL was then added to entirely cover each section. The slides were then placed on the slide rack and inserted into the oven for 30 min at 40°C. The slides were then taken out and slides were washed with 1X Wash Buffer for 2 min at RT twice. AMP 4-FL (Alt A, B, or C) was then added to entirely cover each section. The slides were then placed on the slide rack and inserted into the oven for 15 min at 40°C. The slides were then taken out and slides were washed with 1X Wash Buffer for 2 min at RT twice. The slides were then counterstained with DAPI (30 s at RT) and mounted with ProLong Gold Antifade Mountant (Cat# P10144). We then placed a 24 mm x 50 mm coverslip over the tissue section and stored them in the dark at 4C.

Immunostaining

For SARS-CoV2 spike protein immunostaining, fresh frozen sections were rinsed once with PBS before addition of 10ug/mL of antirabbit IgG Alexa 488 (Invitrogen) + 10ug/mL of SARS-CoV2-spike-RBD protein (Sino Biological) diluted in PBS + 0.5% BSA for 30 mins shielded from light. Two consecutive washes were then performed with PBS + 0.5% BSA 10 min each with gentle agitation before imaging.

For neuro-ectoderm staining, fresh frozen sections were rinsed onced with PBS before fixation at room temperature for 15 min with 4% paraformaldehyde. Three consecutive washes were then performed with PBS 5 min each before addition of blocking buffer (5% normal donkey serum, 0.2% triton x-100 in PBS) for 1 hr. Primary antibody (anti-PAX6 rabbit [Millipore Sigma] diluted 1:50 in blocking buffer) was added overnight (12 hr) at 4C. Three consecutive washes were then performed with PBS 10 min each with gentle agitation before addition of secondary antibody (Anti-Rabbit Dylight 550 (Abcam) diluted 1:200 in blocking buffer) for 1 hr at 37C shielded from light. Three consecutive washes were then performed with PBS 10 min each with gentle agitation before addition of DAPI (1:10,000 dilution in PBS) for 10 min. This was finally followed by three consecutive washes with PBS 10 min each with gentle agitation before imaging.

Microscopy

Following 24 hr of incubation with RNAScope® probes in 4°C, slides were imaged using Zeiss 880 Airyscan Confocal microscope with special thanks to Michael Hu for image processing utilizing the UC San Diego Microscopy Core. Raw images on the Leica DMi8 were obtained with 16bit bit-depth per color, and highlights and shadows were adjusted in the LASX software. Raw images on the Zeiss 880 were obtained with 16bit bit-depth per color, and highlights and shadows were adjusted in the ZEN software. RNAScope images were dilated using ImageJ's MorphoLib by splitting the image into the composite channels and dilating the dots in the appropriate channel. Dots were dilated to 3 pixels as disks.

Cost Analysis

Overall, the cost of profiling a single teratoma with the 10X RNA-seq system runs at about \$1,300, including sequencing costs for \sim 8,000 cells (the output of a single 10X RNA-seq run) at a sequencing depth of 50,000 reads per cell. Mouse husbandry and reagents

CellPress

Cell Resource

related to teratoma formation (cells, Matrigel, media) are relatively cheap in comparison. During teratoma growth, the researcher needs to only monitor the mice for health concerns, weights, and tumor measurements if desired. The teratoma can be extracted at any time after 3 weeks of growth. For the miRNA molecular sculpting experiments the mice require a daily dose of GCV until time of tumor extraction. It is also theoretically possible to inject both flanks of the mouse to generate 2 teratomas per animal. With the availability of easy to use analysis tools such as Seurat/PAGODA2, as well as methods for integrating datasets (such as CONOS), running a basic clustering and cell type annotation of scRNA-seq data is fairly straightforward.

QUANTIFICATION AND STATISTICAL ANALYSIS

Overview

For all figures, we used the CellRanger pipeline as described in the Single Cell RNA-Seq Processing section to generate counts matrices (Zheng et al., 2017). We also used the Seurat package in the R Statistical Programming Language for clustering, data integration, and classification for all figures as described in the Seurat Data Integration and H1 Teratoma Clustering and Validationmethods sections (Stuart et al., 2019). For assigning lentiviral barcodes and CRISPR guide RNAs to cells (relevant to Figures 2 and S2 and Figures 4 and S4 respectively), we used the genotyping-matrices method as described in the Lentiviral Barcode and CRISPR Guide Assignment section (Parekh et al., 2018). For Figures 3 and S3, we used Similarity Weighted Nonnegative Embedding (SWNE) as described in the Developmental Staging Analysis section (Wu et al., 2018). For Figure 4, we quantified guide RNA editing using CRISPResso (Pinello et al., 2016). And for Figure S4, we used DESeq2 as described in the PGP1 Neural Disorder Screen Analysis section (Love et al., 2014). The remaining analysis was done using custom R scripts.

For the heterogeneity analysis in Figures 2 and S2, we treated each teratoma as an individual data replicate. For Figure S4, we collapsed the expression all cells with the same cluster and guide RNA identity into a single replicate in order to run pseudobulk differential expression analysis. For Figure S5, each teratoma was treated as a replicate to compute the cell type proportion z-scores. In other analyses each cell was treated as a replicate.

A brief summary of the analysis details for each figure can be found in the results and figure legends. Below we also provide a mapping between each figure and the relevant methods sections:

- Figures 1 and S1: Seurat Data Integration and H1 Teratoma Clustering and Validation
- Figures 2 and S2: Quantitative Assessment of Teratoma Heterogeneity and Cell Type Bias and Lentiviral Barcode and CRISPR Guide Assignment
- Figures 3 and S3: Developmental Staging Analysis
- Figures 4, S4, and S5: PGP1 Embryonic Lethal Screen Analysis, PGP1 Neural Disorder Screen Analysis and Molecular Sculpting Analysis

All analysis code as well as instructions on how to reproduce our analyses can be found at the Github repository: yanwu2014/teratoma-analysis-code.

Single Cell RNA-seq Processing

Using the 10X Genomics CellRanger (v2.01) pipeline (Zheng et al., 2017), we aligned Fastq files to a combined hg19 and mm10 reference using STAR aligner (Dobin et al., 2013), counted UMIs to generate human and mouse gene-expression counts matrices, and aggregated samples across 10X runs with the cellranger aggr command. All cellranger commands were run using default settings.

Seurat Data Integration

Data integration was performed on the aggregated counts matrices for each of the following datasets: the 7 H1 teratomas, the 6 PGP1 CRISPR-KO screen teratomas, and the 3 cell line teratomas. We used the Seurat v3 data integration pipeline (Butler et al., 2018; Stuart et al., 2019). Briefly, we first filtered the counts matrix for genes that are expressed in at least 0.1% of cells, and cells that express at least 200 genes. We then normalized the counts matrix using total-counts normalization, and log-transformed the result. Log-transforming RNA-seq counts results in the data following an approximately normal distribution, which is the assumption that Seurat makes for the remainder of the analysis (Law et al., 2014). For each teratoma, we identified highly variable genes, and selected the top 4000 genes that appeared as overdispersed across the most teratomas. We then identified anchor cells, and integrated the teratomas to create a batch-corrected gene expression matrix. After batch correction, we used a linear model to regress away library depth, and mitochondrial gene fraction, and ran Principal Components Analysis (PCA) (Abdi and Williams, 2010), keeping the first 30 principal components. We then used the PCs to generate a k Nearest Neighbors (kNN) graph, setting k = 10, and then used the kNN graph to calculate a shared nearest neighbors (SNN) graph (Houle et al., 2010). We ran modularity optimization algorithm with a resolution of 0.4 on the SNN graph to find clusters (Butler et al., 2018).

H1 Teratoma Clustering and Validation

H1 clusters were assigned to cell types using a two-stage strategy. First, we trained a kNN classifier on the Mouse Cell Atlas dataset using k = 40 (Tarlow et al., 2013), mapping mouse genes to their human orthologs. We projected each cell in the teratoma dataset onto





the first 40 Principal Components (PCs) of the Mouse Cell Atlas and classified each cell in the H1 teratoma dataset using this kNN classifier to generate a rough set of cell type assignments for each cluster. We then manually inspected the marker genes for each cluster and adjusted the cell type based on the expression of canonical markers (Tables S2A–S2E). We also specifically looked at transcription factor markers using the TRRUST database (Tables S2A and S2D) (insert reference). We computed differential gene expression in Seurat using the default Wilcoxon rank-sum test, which does not make any assumptions about the distribution of the data being tested, otherwise known as a non-parametric test (Wilcoxon, 1946). Clusters that mapped to the same MCA cell type, and expressed similar marker genes were merged. Finally, we ran UMAP on the first 30 PCs as input in order to visualize the results (Becht et al., 2018; McInnes and Healy, 2018). We validated each annotated cell type by computing the Pearson correlation between the average expression of each cell type and the average expression of each broad cell type in the Mouse Organogenesis Cell Atlas (Cao et al., 2019). We used the union of all marker genes for the teratoma cell types and Mouse Organogenesis Cell Atlas cell types to perform the correlation analysis.

In some cases, it was necessary to sub-cluster the cells to achieve greater cell type resolution. Specifically, we noted that the ciliated epithelium cluster had both retinal and airway markers so we sub-clustered the all cells mapping to ciliated epithelium in order to separate retinal epithelium and airway epithelium. Additionally, we sub-clustered the neuro-ectoderm in order to identify interneurons, peripheral neurons, retinal progenitors, and early neuro-ectoderm. In both cases we simply subsetted the gene expression matrix with the cells of interest and reran the Seurat analysis pipeline, identifying sub-clusters using known marker genes (Table S2G).

Quantitative Assessment of Teratoma Heterogeneity and Cell Type Bias

In order to quantify the level of heterogeneity between teratomas we used the Normalized Relative Entropy metric from CONOS (Barkas et al., 2019).

$$1 - \frac{\sum_{k=1}^{n_{clust}} \mathbf{s}_k \times KL(f_k, F)}{\log(n_{teratomas}) \sum_{k=1}^{n_{clust}} \mathbf{s}_k}$$

Where f_k is a vector with the number of cells in each teratoma from cluster k, $KL(f_k, F)$ is the empirical KL divergence between f_k and the total number of cells in each teratoma, F. Higher Normalized Relative Entropy means the cell types are more mixed across the teratomas and thus the teratomas are less heterogeneous.

There was only one replicate per non-H1 cell line teratoma as our main goal was to assess the heterogeneity across cell lines versus the heterogeneity within the H1 cell line, while also demonstrating that we could generate teratomas using multiple cell lines.

To quantify the heterogeneity/bias of individual cell types across teratomas we simply take the KL divergence of the number of cells in each teratoma from that cell type/cluster and the total number of cells in each teratoma and then scale by the number of cells in each cell type. For each cell type *k*:

$$s_k \times KL(f_k, F)$$

Lentiviral Barcode and CRISPR Guide Assignment

To assign one or more lentiviral/gRNA barcode to each cell, we extracted each barcode by identifying its flanking sequences, resulting in reads that contain cell, UMI, and barcode tags. To remove potential chimeric reads, we used a two-step filtering process. First, we only kept barcodes that made up at least 0.5% of the total amount of reads for each cell. We then counted the number of UMIs and reads for each plasmid barcode within each cell, and only assigned that cell any barcode that contained at least 10% of the cell's read and UMI counts. The code for assigning barcodes to each cell can be found on GitHub at: https://github.com/yanwu2014/ genotyping-matrices (Parekh et al., 2018).

H1 Cell Barcoding Analysis

We extracted lentiviral barcodes from the genomic DNA fastq files before and after teratoma formation for the 3 barcoded H1 teratomas. We counted the number of unique barcodes that were supported by at least 10 reads (the reads requirement is to mitigate overcounting unique barcodes due to minor sequencing errors) and then computed the fraction of unique barcodes that remain after teratoma formation to assess the approximate number of cells that are involved in the teratoma formation process.

We also identified lentiviral barcodes at the single cell level, using the barcode assignment strategy described in the Lentiviral Barcode and CRISPR Guide Assignment section. For each cell type, we computed its bias for specific barcodes using the same relative entropy metric we used to compute teratoma bias.

$$s_k \times KL(b_k, B)$$

Where b_k is a vector with the number of cells in each barcode from cluster k, $KL(b_k, B)$ is the empirical KL divergence between b_k and the total number of cells in each barcode, B.

Developmental Staging Analysis

In order to assess the developmental maturity of the teratoma cell types, we computed the average expression of all cells related to neuro-ectoderm (Radial Glia, Intermediate Neuronal Progenitors, Early Neurons) and gut (Oral/Esophageal, Stomach, Intestine) cell

CellPress



types and calculated the cosine similarity of the teratoma average expression to the average expression of fetal human cells across different time points. We used all genes that were detected in both the fetal and teratoma data.

For the neuro-ectoderm cells, we then sub-clustered those cells and identified additional cell types using canonical marker genes (Table S2G). We then matched those neuro-ectoderm sub-clustered cell types to cell types in a larger fetal week 17-18 single cell prefrontal cortex dataset.

We next generated Similarity Weighted Nonnegative Embeddings (SWNE) (Wu, Tamayo and Zhang, 2018a) for the neuronal and gut cell types using the top 3000 overdispersed genes in each tissue type. Briefly, SWNE uses nonnegative matrix factorization (NMF) (Lee and Seung, 1999) to decompose a gene expression matrix into component factors, embeds the factors in 2D using sammon mapping (Sammon, 1969), and embeds the cells and key genes in the 2D space relative to the factors. The cell positions are smoothed using a shared nearest neighbors (SNN) network. For the neuronal SWNE embedding, we used 30 NMF factors and 20 nearest neighbors when computing the SNN. For the gut SWNE embedding, we used 20 NMF factors and 30 nearest neighbors. We projected teratoma data onto the fetal SWNE, by first projecting the teratoma data onto the fetal SNN.

We then compared the expression of key neuronal/gut marker genes in each neuronal and gut cell type by correlating the expression of those markers between the teratoma data and the fetal human data. We used the scaled gene expression for both the teratoma and fetal data, which involves subtracting the average expression and dividing by the standard deviation. We selected the cell type markers for the neuro-ectoderm and gut comparisons using published studies of the developing human cortex and developing gut. Specifically, we selected VIM/SOX2 as markers for Radial Glia, DLX1 as a marker for Interneurons, and HMGB2 as a marker for Cycling Progenitors using the markers from the single-cell RNA-seq study of week 17 – 18 developing human cortex (Polioudakis et al., 2019). HES5 is known to be a key regulator of the neural progenitor state while DCX and NEUROD1 are essential for early neuronal differentiation (Gao et al., 2009b; Bansod et al., 2017; Khalaf-Nazzal et al., 2017). For the developing gut, we selected CDX1/CDX2 as Mid/Hindgut markers and PAX9 as a foregut marker from the single-cell RNA-seq study of the developing human digestive tract (Gao et al., 2018). HHEX regulates midgut development, specifically the formation of the pancreas from the gut tube (Bort et al., 2004). SOX2 is a known foregut marker that regulates gut patterning while FOXJ1 marks foregut cells primed for the lung epithelial lineage (Que et al., 2007; Green et al., 2011).

PGP1 Embryonic Lethal Screen Analysis

For each of the six teratomas across the original and replicate screens, we used two technical replicate 10X runs. In order to ensure consistent cell types across teratomas, we merged the 10X runs corresponding to the same teratoma, and then integrated all six teratomas across both the original and replicate screen using Seurat v3 data integration. We used 3000 anchor features and 20 CCA dimensions for the integration. Using the annotated H1 teratoma dataset as the reference, we used Seurat label transfer to identify the cell type for all cells in the screen datasets. Due to the relatively low number of cells per guide RNA in the original screen, we collapsed closely related cell types into broader cell groupings in order to boost the power of our analysis. Specifically, Airway Epithelium was merged into Foregut (Airway epithelium is derived from the foregut epithelium during development), Schwann Cells and Melanoblasts were grouped as Schwann Cell Progenitors (SCP), Immune Cells, Erythrocytes, and Hematopoietic Stem Cells (HSCs) were grouped as Hematopoietic cells, Muscle Progenitors (INP) and Radial Glia were grouped as Neuronal Progenitors, and Retinal Neurons and Early Neurons were simply grouped as Neurons. In order to visualize the PGP1 data, we projected the integrated screen dataset onto the first 20 PCs from the H1 dataset and ran UMAP on the projected PCs.

We validated the editing efficiencies of all our guide RNAs using PCR amplification of the expected cut site and looking for mutations and indels with CRISPResso. We then selected the top guide targeting each gene with at least a 60% overall editing efficiency and a 40% indel efficiency which resulted in a total of 16 out of 48 guides selected. We then only used these 16 validated guides for further computational analysis. Unfortunately, the TULP3-2 guide was not detected in the replicate screen so we ended up using 15 guides (plus 5 NTC guides) for analysis.

We assigned CRISPR-KO gene perturbations using the barcode assignment strategy described in the Lentiviral Barcode and CRISPR Guide Assignment section. To determine the total effect of each knockout, we computed a normalized Earth Mover's Distance (EMD) between all cells in each gene knockout with all cells belonging to the NTC separately for each screen (Chen et al., 2020). EMD computes the difference in cell type composition between two groups of cells, weighted by how transcriptionally distinct the cell types are (Chen et al., 2020). Thus, differences in cell type composition between cells belonging to the gene knockouts and NTC that arise from the fact that the label transfer has a hard time distinguishing similar cell types will not be as highly weighted as differences between distinct cell types. We ran the EMD analysis separately for the original and replicate screens, and normalized the EMD metric so that the average EMD for all NTC guides would equal 1.

To assess the effect of gene knockouts on individual cell types, we used a ridge regression model with the R glmnet package as initially described in the PerturbSeq method (Friedman et al., 2015; Dixit et al., 2016). Briefly, for each CRISPR gRNA, this resulted in regression coefficients for each cell type describing the enrichment or depletion of that gRNA in that cell type. This method assumes that the data is normally distributed, which is approximately true for RNA-seq and scRNA-seq data when log-transformed (insert ref). We permuted the gRNA assignments to assign p values to each coefficient representing the probability that coefficient is non-zero by chance. Because we used a non-parametric permutation test, we did not make any assumptions about the distribution of regression





coefficients. We then used the Benjamini-Hochberg multiple testing correction (Thissen et al., 2002) to generate False Discovery Rates and visualized coefficients with an FDR < 0.05. For each gRNA, we computed the cell type shift effect size as the average EMD effect across the screens. The reproducibility of the gRNA knockout was assessed by correlating the gRNA knockout effects (regression coefficients) across the original and replicate screen.

PGP1 Neural Disorder Screen Analysis

For each of the 2 teratomas across the original and replicate screens, we used two technical replicate 10X runs. In order to ensure consistent cell types across teratomas, we merged the 10X runs corresponding to the same teratoma, and then integrated the teratomas using Seurat v3 data integration. We used the same data integration and label transfer parameters as the embryonic lethal screen. We again collapsed closely related cell types into the broader cell groupings described in the PGP1 Embryonic Lethal Screen section, and additionally filtered out any remaining cell types with fewer than 200 cells.

We assigned CRISPR-KO gene perturbations using the barcode assignment strategy described in the Lentiviral Barcode and CRISPR Guide Assignment section. To determine the total effect of each knockout, we again computed the normalized Earth Mover's Distance (EMD) between all cells in each gene knockout with all cells belonging to the NTC separately for each screen (Chen et al., 2020).

We analyzed differential expression for each broad cell type separately so that cell type specific effects would be captured. For each cell type, we summed the counts for all cells assigned to a specific guide RNA and a specific teratoma to create a pseudobulk expression matrix. This essentially treats each guide in each teratoma as a biological replicate for a given gene knockout, and enables us to use DESeq2, a well-validated differential expression method (Love et al., 2014). For each gene knockout, we ended up with 6 pseudobulk replicates (3 guides x 2 teratomas). We ran DESeq2 with default parameters, comparing the pseudobulk replicates for each gene with the NTC replicates, and used apeglm to shrink effect sizes. We set a False Discovery Rate cutoff of 0.1 to call a gene differentially expressed. We also ran DESeq2 on each teratoma separately to compute log fold-changes and assess reproducibility.

Molecular Sculpting Analysis

To assess the enrichment or depletion of cell types in the miRNA-HSV-tk transduced H1 teratomas, we compared teratomas that had ganciclovir (GCV) added using intratumoral (IT) and both intratumoral and intraperitoneal (IPIT) injection methods, versus a control teratoma that had the construct miRNA-HSV-tk but no GCV. All teratomas were injected on the same date and extracted after 10 weeks of growth. To assign cell types, we again used Seurat's label transfer. We then collapsed cell types using the same merging strategy described in the PGP1 Teratoma Screen Analysis section, and then computed the fraction of cell types present in each teratoma. Finally, we computed log2 fold-changes of cell type fractions by dividing the cell type fractions in the GCV+ IT/IPIT teratomas with the cell type fractions in the GCV- teratoma. To compute an estimated z-score, we subtracted the GCV- teratoma fractions from the GCV+ IPIT/IT teratoma fractions and divided by the cell type fraction variance. The z-scores for IPIT and IT teratomas were computed separately, and the cell type fraction variance was computed by pooling the variance of the miRNA-HSV-tk teratomas and the variance of the plain H1 teratomas with Cohen's pooled standard deviation (Cohen, 1988).

Figure Generation

All figures were generated using original artwork or open source with InkScape, Adobe Illustrator®, and ImageJ.



Cell Resource

Supplemental Figures





(A) H&E stains (left to right, top to bottom): Choroid Plexus, Fetal Neuro-ectoderm, Retinal Pigment Epithelium (RPE), Developing Airway, Ciliated Respiratory Epithelium, Fetal Cartilage, Mesenchyme, Bone, Developing Cardiac/Skeletal Muscle, Squamous epithelium, Retinal Neurons (around RPE), Smooth Muscle, Adipocytes. (B) The fraction of cells that are classified as MSC/Fibroblast across each teratoma. (C) Heatmap of the average expression of key marker genes for each cell type (guidelines separate cell types from different germ layers) (Table S3C). (D) Correlation of the average expression of each human teratoma cell type with the average expression of each fetal mouse cell type. (E) UMAP plot of mouse cell types in the H1 teratomas.







Figure S2. Assaying Teratoma Heterogeneity, Related to Figure 2

(A) UMAP scatterplot showing how each line (HUES62, PGP1, and H9) contributes to the various cell type clusters. (B) Left: the normalized proportion of each teratoma in every cell type. Right: the bias each cell type shows toward specific teratomas. A low bias score means the cell type is well mixed across all 7 teratomas. (C) Growth kinetics of 6 teratomas based on cell line (HUES62, PGP1, and H9). (D) Karyotyping of all 4 PSC lines. (E) Lentiviral barcode construct map. (F) Barcoding summary statistics for both bulk and single cell assays across the three barcoded teratomas.

Cell Resource

CellPress



(legend on next page)





Figure S3. Assaying Teratoma Maturity, Related to Figure 3 and Table 1

(A) A heatmap of log fold-changes for the top differentially expressed genes between matched teratoma neuro-ectoderm and fetal cortical cell types. (B) A heatmap of the enrichment scores for top differential genesets (via Geneset Enrichment Analysis) between matched teratoma neuro-ectoderm and fetal cortical cell types. (C) Cosine similarity of teratoma gut cells with fetal gut cells of different ages. (D) Projection of fetal gut epithelium cell types onto a teratoma gut epithelium SWNE embedding. (E) Correlation of the scaled expression of key marker genes across mid/hindgut epithelium and foregut epithelium between teratoma and fetal cell types. Marker genes are colored by their corresponding cell type. (F) Proportion of foregut and mid/hindgut cells in the teratoma and fetal gut. (G) A heatmap of log fold-changes for the top differential genesets (via Geneset Enrichment Analysis) between matched teratoma gut epithelium and fetal gut epithelium cell types. (H) A heatmap of the enrichment scores for top differential genesets (via Geneset Enrichment Analysis) between matched teratoma gut epithelium and fetal gut epithelium cell types. (H) A heatmap of the enrichment scores for top differential genesets (via Geneset Enrichment Analysis) between matched teratoma gut epithelium and fetal gut epithelium cell types. (H) A heatmap of the enrichment scores for top differential genesets (via Geneset Enrichment Analysis) between matched teratoma gut epithelium and fetal gut epithelium cell types. (I) H&E stains (left) as well as RNA FISH staining (right) of FOXJ1 (Airway epithelium), CDX2 (Intestinal epithelium), TNNT2 (Cardiac muscle), and THY1 (mesenchymal stem cell/fibroblast). Scalebar = 50μM (20x). Dots were dilated using ImageJ.







Figure S4. Engineering Teratomas via Genetic Perturbations, Related to Figure 4

(A) Schematic showing knock-in of the CAG-spCas9-P2A-EGFP cassette with an upstream T2A linked blasticidin resistance gene into the AAVS1 locus thus, creating the Cas9-expressing PGP1 line (above). Accompanying validated trace sequences of the left and right arms (below). (B) 2% agarose gel confirming integration of the CAG-spCas9-P2A-EGFP cassette into the AAVS1 locus of the PGP1 line via PCR amplification of the left and right arm spanning the endogenous locus and the engineered cassette compared to a PGP1 negative control. (C) Observed cells per gRNA and cells per gene for the screen. (D) UMAP projection of PGP1 cell types classified using the H1 cell types as a reference. (E) PGP1-Cas9 iPSCs were transduced with a CRISPR-Cas9 library targeting *TCF4* (Pitt-Hopkins Syndrome), *MECP2* (Rett Syndrome), and *L1CAM* (L1 Syndrome) with 3 guides each. After generating 2 teratomas with the PGP1-iPSCs, scRNA-seq was used to identify shifts in cell type specific gene expression as a result of gene knockouts. (F) Shift in cell types a measured by normalized Earth Mover's Distance (EMD) due to knockouts from the embryonic lethal knockouts and the disease screen knockouts (*TCF4, MECP2, L1CAM*). (G - I) The shift in gene expression as measured by log2 fold-change against NTC guides across both teratoma replicates for (G) L1CAM knockout in Neurons, (H) MECP2 knockout in Neural Progenitors. The color of the data points represents the –log(False Discovery Rate) as computed by DESeq2.



2 m

Cell **Resource**







Figure S5. Engineering Teratomas via Molecular Sculpting, Related to Figure 4

(A) Phase images from light microscopy showing H1 cell survival after 3 and 5 days in the presence of GCV (10μM). H1 ESC line was either transduced with GFP control (EGIP backbone) or miR-124-HSV-tk-GFP. (B)-(C) Quantification using flow cytometry and gating based on the presence or absence of GFP in HEK293T (red) and HeLa/HUVEC (blue) cells (B)/(C) transduced with either No GFP control, HSV-tk-GFP, or miR-21-HSV-tk-GFP/miR-126-HSV-tk-GFP for 5 days (STAR Methods). (D) Schematic of generating self-patterned whole brain organoids (STAR Methods). (E) Images of teratomas grown in the absence and presence of GCV administration (80mg/kg/d, STAR Methods) for 10 weeks. (F) H&E stains of teratomas grown in the absence (left) and presence (right) of GCV administration. Arrowheads highlight regions of neuro-ecotoderm. Scalebars are directly labeled. (G) anti-*PAX6* (red) and DAPI (blue) immunostaining in GCV+ and GCV- control sections across 3 different regions of the corresponding teratoma. Scalebar = 2 mm. (H) Secondary antibody staining only (Dylight 550, red) and DAPI (blue) for a GCV+ and GCV- teratoma. Scalebar = 2 mm. (I) RNA FISH analysis of *HES5* (red) and DAPI (blue) in a GCV+ and GCV- teratoma. Scalebar = 2 mm, 200 μm (magnified insert).