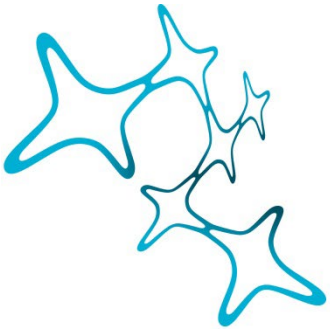
Resolving reprogramming factor and starter cell impact on cell fate conversion

Bobby Aron Hersbach



Graduate School of Systemic Neurosciences

LMU Munich

Dissertation der

Graduate School of Systemic Neurosciences

Ludwig-Maximilians-Universität München

01.03.2022

Supervisor

Prof. Dr. Magdalena Goetz

Department of Physiological Genomics

Ludwig-Maximilians-Universität München

First Reviewer: Prof.Dr. Magdalena Götz

Second Reviewer: Prof.Dr. Stefan Stricker

External Reviewer Prof.Dr. Malin Parmar

Date of Submission: 01.03.2022

Date of Defense: 23.06.2022

CONTENTS

SUMMARY

The advent of cellular reprogramming strategies has been one of the major breakthroughs in the field of biology, holding great promise for disease modelling and regenerative medicine. However, although the technique has celebrated numerous successes over the past decades, our fundamental understanding of its mechanisms remains remarkably limited. During my PhD, I have focused on unravelling and understanding some of these mechanisms, with the final aim of improving current reprogramming strategies. To this end, my work was divided into two complementary projects that each concentrated on a different aspect of reprogramming: the reprogramming transcription factor and the starter cell.

In the first part, we compared different reprogramming factors in their ability to shut down the original cell identity and assessed to what extent their expression levels are of importance to conversion. Additionally, we pitted different reprogramming factors against each other to unravel whether hierarchies exist when reprogramming to more than one identity. These experiments revealed a general pattern where the erasure of the original identity as well as the establishment of a new identity was most efficient when performed within a single germ layer. In addition, a general threshold of reprogramming factor expression levels was found, with additional increases in reprogramming factor expression having little effect on conversion. Upon collision, the transcription factor rather than germ layer affiliation was decisive, possibly influenced by the availability of cofactors. Furthermore, insensitivity of these collisions to relative reprogramming factor expression levels revealed general determinism of cell states as a function of these expression levels, facilitating model-based factor selection.

In the second part, we assessed the influence of starter cell identity on direct neuronal conversion. Specifically, we concentrated on astrocyte to neuron conversion using astrocytes obtained from distinct brain regions. The use of such regionally heterogenous astrocytes revealed that the actions of the same reprogramming transcription factors are highly specific to the cellular environment. Moreover, the actions of reprogramming factors are shaped in such a way that they eventually steer towards developmentally and regionally appropriate neuronal subtypes.

Taken together, these results provide several key insights into the mechanisms of reprogramming. Most importantly, they warrant a more holistic view of the reprogramming process where germ layer identity, reprogramming factor expression levels, starter cell identity and model-aided factor selection should all be considered when designing and improving reprogramming paradigms.

INTRODUCTION

Establishment of cell identity

Mammalian development is an intricate process that ultimately generates the vast range of different cell types that make up the adult body. For a long time, one of the major challenges in the field of biology has been to determine how this process is orchestrated, i.e., how a single fertilized egg generates this remarkable diversity. In 1957, the developmental biologist Conrad Waddington proposed a conceptual visualization of this process (Waddington, 1957). In his *Epigenetic Landscape*, Waddington depicted a differentiating cell as a marble rolling down a landscape of valleys and hills (**Figure 1A**). As the marble progresses downwards, the landscape bifurcates, representing distinct cell fate decisions (**Figure 1A**). Eventually, the marble comes to rest in one of the many valleys at the bottom of the landscape, corresponding to the acquisition of a terminally differentiated cell state (**Figure 1A**). Waddington considered this to be a unidirectional process, illustrated by the hills that separate the valleys from each other. Furthermore, he envisioned a network of pegs and strings shaping the landscape from underneath to guide these cell fate decisions, although he did not know their biological counterpart at this stage (**Figure 1B**).

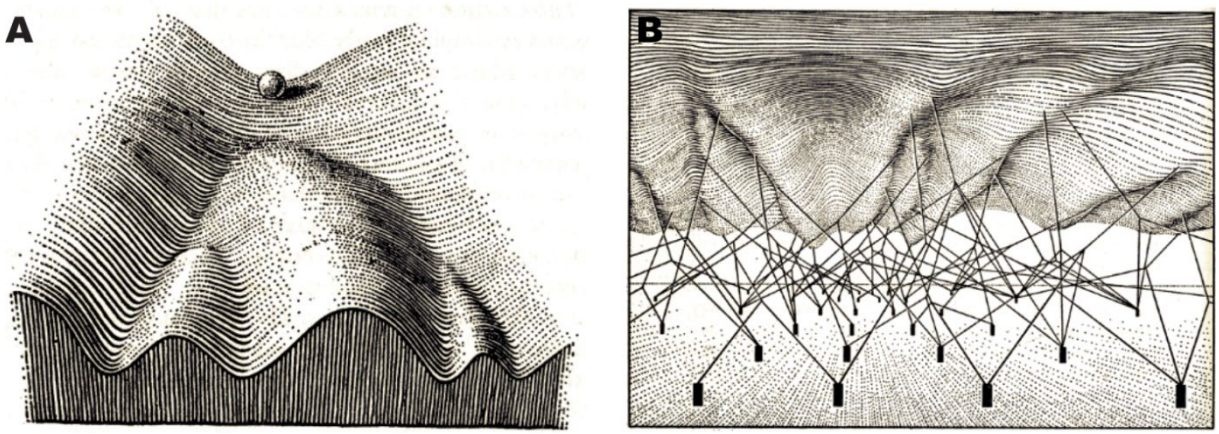


Figure 1: Waddington's epigenetic landscape. A. Surface of the landscape depicting the development paths available to a marble that represents a differentiating cell **B.** The pegs and strings underlying the landscape shaping cell fate decisions. Adapted from Moris, Pina and Arias, 2016 with permission from Springer Nature (License No. 5235400697582).

Although highly simplified, this model turned out to be an effective way to depict some of the key concepts of cell identity establishment during development (**Figure 2**). At the top of the hill is the fertilized egg, or zygote, the totipotent cell that will generate all cells that make up the adult organism (Condic, 2014). As development commences (i.e., the marble starts to roll down

the hill) diversification of cell identity is achieved through the establishment of distinct gene expression patterns which drive discrete developmental trajectories, thereby representing the pegs and strings depicted by Waddington (Reviewed by Zernicka-Goetz, Morris and Bruce, 2009). Differences in gene expression patterns come about by the differential activity of transcription factors, a class of proteins that interacts with the DNA to initiate transcription. In turn, variation in transcription factor activity results from their capacity to integrate several extrinsic cues such as cellular location, polarity and orientation of the division plane. For example, the first bifurcation in the landscape, and thereby the first cell fate decision, occurs when distinct transcriptional programs between cells are established as a result of differences in cellular position within the embryo and set apart the pluripotent inner cell mass (ICM) from the trophectoderm (TE) (**Figure 2**)(Barlow, Owen and Graham, 1972; Johnson and Ziomek, 1981; Pedersen, Wu and BaŁakier, 1986; Dyce *et al.*, 1987; Fleming, 1987). Shortly thereafter, cellular interactions between cells within the ICM result in the second cell fate decision by establishing two different transcriptional programs that define the primitive endoderm (PE) and the epiblast (Epi)(Gardner, 1982; Chazaud *et al.*, 2006; Plusa *et al.*, 2008). In summary, the actions of transcription factors, translating cell extrinsic cues into cell intrinsic gene expression patterns, shape the landscape that gives rise to the lineages of the early embryo (**Figure 2**).

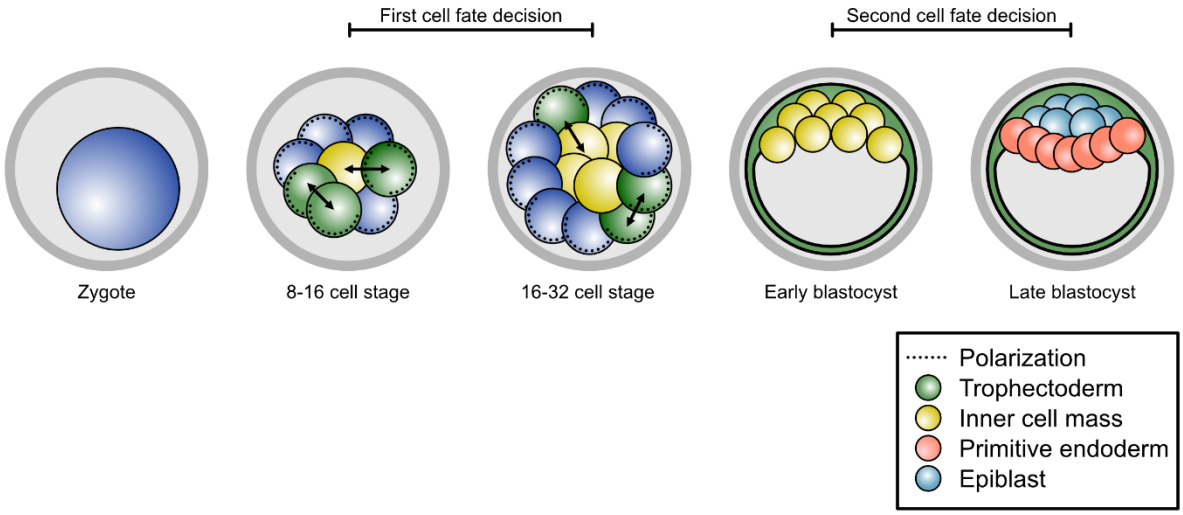


Figure 2: Early mouse embryogenesis. Schematic overview of the initial stages of mouse embryonic development including the first (Inner cell mass and Trophectoderm) and second (Primitive endoderm and Epiblast) cell decision.

In the next phase, the newly generated epiblast cells will give rise to three distinct populations of lineage biased progenitor cells that will each give rise to distinct tissues and organs (see below). The start of this process, commonly referred to as gastrulation, is marked by the appearance of the primitive streak, a visible groove in the layer of epiblast cells on one side of the embryo (**Figure 3**). Subsequently, epiblast cells start to detach from one another and

migrate towards the primitive streak where they ingress and form a new layer of cells: the endoderm (Lawson, Meneses and Pedersen, 1986; Lawson and Schoenwolf, 2003; Kimura *et al.*, 2006). In the next wave of migration, cells form a second layer in between the newly formed endoderm and the remaining epiblast cells, referred to as the mesoderm (Ferretti and Hadjantonakis, 2019). Finally, epiblast cells that fail to migrate through the primitive streak will form the last germ layer, the ectoderm (Lawson and Pedersen, 1992). Throughout this process, cellular location and timing of migration determine the unique combination of signaling cues a cell is exposed to, which is translated into distinct transcription factor expression patterns that drive cell fate decisions and result in the distinct identity of the three germ layers. However, although cells from each germ layer are biased towards a certain lineage, they are still multipotent, meaning that they will need to undergo several further lineage decisions to determine their final identity (**Figure 1A**). Below, we will briefly provide some examples of such lineage decisions for all three germ layers and highlight a few of the transcription factors involved.

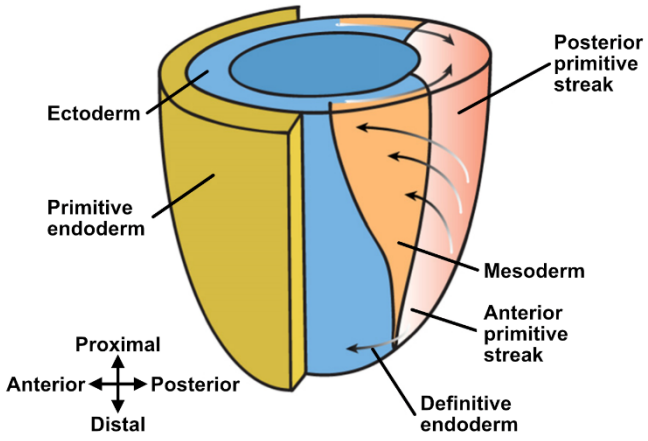


Figure 3: Mouse gastrulation. Schematic illustration of mouse gastrulation depicting the primitive streak (pink) where epiblast cells (blue) ingress to form the endoderm (yellow), mesoderm (orange) and ectoderm (blue). Adapted from Duelen and Sampaolesi, 2017 under the Creative Commons License v4.0.

Endoderm

Cells from the endoderm give rise to the gut as well as several other visceral organs including the lungs, pancreas, and liver. Remarkably, development of the liver requires only a small number of transcription factors that commit multipotent endoderm cells to liver progenitors: Forkhead box protein A (FoxA) 1 and 2 (Lee *et al.*, 2005) as well as GATA-binding proteins (Gata) 4 and 6 (Holtzinger and Evans, 2005; Zhao *et al.*, 2005; Watt *et al.*, 2007; Si-Tayeb, Lemaigre and Duncan, 2010). Indeed, when both FoxA1 and FoxA2 are lacking, liver formation

is completely absent (Lee *et al.*, 2005). However, FoxA factors only seem to be important for the initial specification of liver progenitors, as they are no longer required for subsequent differentiation stages (Li *et al.*, 2009). Instead, a network of hepatocyte nuclear factors (Hnfs) drives further liver differentiation by establishing hepatic gene regulatory networks (Odom *et al.*, 2004). In short, once liver progenitors are specified, the expression of One cut homeobox 1 (Onecut1) induces the expression of Hnf1β, which establishes initial hepatic gene expression (Cereghini *et al.*, 1992; Lazzaro *et al.*, 1992; Clotman *et al.*, 2002). Next, Hnf1β and Gata6 together drive the expression of Hnf4α, one of the central regulators of hepatic transcriptional programs and key to the final maturation of liver cells (Kuo *et al.*, 1992; Hatzis and Talianidis, 2001; Watt, Garrison and Duncan, 2003; Briançon *et al.*, 2004; Babeu and Boudreau, 2014; Lau *et al.*, 2018).

Mesoderm

Mesoderm cells generate a plethora of different cell types that make up most of the adult body (Ferretti and Hadjantonakis, 2019). These include the cardiovascular system, lymphatic system as well as both smooth and skeletal muscle (Dudek, 2009). Among these derivates, the terminal differentiation of skeletal muscle has been particularly well defined. First, mesodermal stem cells in the embryo form a series of block-like structures referred to as somites (Summerbell and Rigby, 1999; Pourquié, 2001; Pownall, Gustafsson and Emerson, 2002). From these somites, myogenic precursors expressing the transcription factors paired box protein (Pax) 3 and 7 emerge (Jostes, Walther and Gruss, 1990; Goulding, Lumsden and Paquette, 1994; Hirsinger *et al.*, 2001; Kiefer and Hauschka, 2001; Parker, Seale and Rudnicki, 2003). Within these precursor cells, the expression of Pax3 leads to induction of a panel of myogenic regulating factors (MRFs) including myogenic factor (Myf) 5 and 6, myogenic differentiation 1 (MyoD1) and myogenin (Myog), which collectively drive muscle differentiation (Parker, Seale and Rudnicki, 2003). Importantly, despite the fact that all MRFs belong to the same family of basic helix-loop-helix (bHLH) transcription factors, little redundancy exists between them (Hasty *et al.*, 1993; Nabeshima *et al.*, 1993; Braun and Arnold, 1995; Patapoutian *et al.*, 1995; Zhang, Behringer and Olson, 1995; Kablar *et al.*, 1999; Parker, Seale and Rudnicki, 2003). For example, the activation of either MyoD1 or Myf5 is both necessary and sufficient for initial myogenic specification (Braun *et al.*, 1992; Rudnicki *et al.*, 1992, 1993), as double knock-out animals for these factors fail to develop skeletal muscle (Kablar *et al.*, 1999). Myog, on the other hand, is essential for the differentiation and fusion of myoblasts into multinucleated myotubes, as null mice fail to develop adequate muscle fibers while myoblast numbers are unaltered (Hasty *et al.*, 1993; Nabeshima *et al.*, 1993). Finally, Myf6 controls terminal differentiation of skeletal muscle and is

involved in their maintenance (Braun and Arnold, 1995; Patapoutian *et al.*, 1995; Zhang, Behringer and Olson, 1995).

Ectoderm

The ectoderm develops from epiblast cells that fail to migrate through the primitive streak. Once formed, it segregates into the neuroectoderm, which will give rise to the central and peripheral nervous system (CNS/PNS), and the non-neural ectoderm, which will form the epidermis (Lawson and Pedersen, 1992; Lawson, 1999; Patthey and Gunhaga, 2014). The formation of the neuroectoderm is initiated when signaling cues from underlying mesoderm structures cause the ectoderm to thicken and form a structure referred to as the neural plate (Gilbert, 2000). Subsequently, the neural plate folds to form the neural tube which spans along most of the embryo's rostral-caudal axis (Gilbert, 2000). Once again, cellular location determines the unique combination of signaling cues progenitor cells are exposed to and the different gene expression programs that these impose divide the neural tube into different segments that will give rise to distinct structures within the CNS (**Figure 4**) (Cohen, Briscoe and Blassberg, 2013; Bier and de Robertis, 2015). Strikingly, the induction of the neural fate within these segments seems to be driven by a small group of transcription factors, often referred to as proneural genes (Bertrand, Castro and Guillemot, 2002; Baker and Brown, 2018). Proneural genes were initially discovered in *Drosophila melanogaster,* where their expression was found to be sufficient to specify neurogenic progenitor cells from unspecified ectoderm (Knust and Campos-Ortega, 1989; Simpson, 1990; Cubas *et al.*, 1991; García-Bellido and de Celis, 2009; Baker and Brown, 2018). Furthermore, similarities in their sequences led to the identification of the shared bHLH domain, which gives these factors their DNA binding capability (Murre, McCaw and Baltimore, 1989). Homologs in vertebrates were found to have similar proneural capacity (Guillemot *et al.*, 1993; Fode *et al.*, 1998, 2000; Ma *et al.*, 1998, 1999; Casarosa, Fode and Guillemot, 1999; Horton *et al.*, 1999; Scardigli *et al.*, 2001; Cau, Casarosa and Guillemot, 2002). In particular, Achaete-scute homolog 1 (Ascl1) and Neurogenin 2 (Neurog2) were found to be exceptionally potent in inducing inhibitory and excitatory neuronal phenotypes respectively and are both necessary and sufficient to do so in large parts of the developing fetal brain and spinal cord (Casarosa, Fode and Guillemot, 1999; Horton *et al.*, 1999; Fode *et al.*, 2000; Gowan *et al.*, 2001; Nieto *et al.*, 2001). However, their expression is transient and becomes downregulated before progenitor cells commence differentiation (Ben-Arie *et al.*, 1996; Gradwohl, Fode and Guillemot, 1996; Ma, Kintner and Anderson, 1996). Hence, a different set of genes is required for further neuronal differentiation. Indeed, both Ascl1 and Neurog2 induce the expression of several members of the NeuroD family of bHLH factors, which guide the further neuronal differentiation into a variety of different neuronal subtypes (Lee *et al.*, 1995; Ma, Kintner and

Anderson, 1996; Fode *et al.*, 1998; Ma *et al.*, 1998; Miyata, Maeda and Lee, 1999; Perron *et al.*, 1999; Farah *et al.*, 2000; Huang *et al.*, 2000; Liu *et al.*, 2000; Schwab *et al.*, 2000; Olson *et al.*, 2001; Cau, Casarosa and Guillemot, 2002).

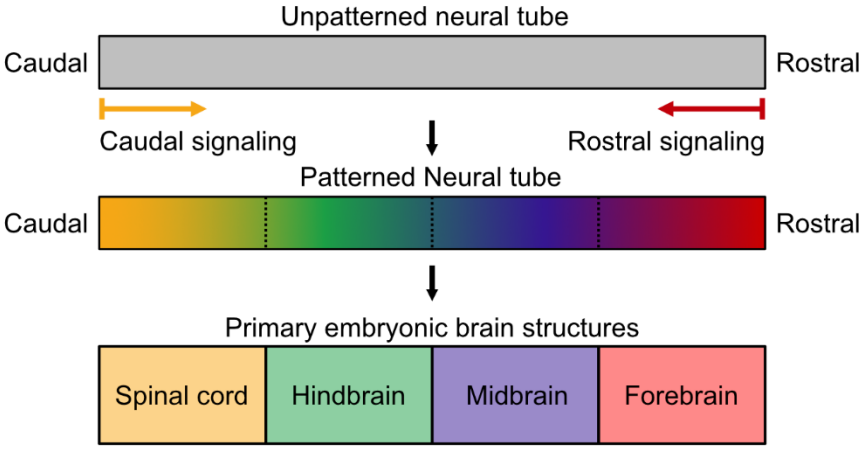


Figure 4: Neural tube segmentation. Diagram depicting the different segments of the neural tube that arise through the unique combination of patterning cues progenitor cells are exposed to along the rostral-caudal axis of the neural tube,

Taken together, the above examples of progenitor differentiation reveal a general pattern across germ layers, where a hierarchical cascade of transcription factors progressively drives differentiation into terminally differentiated cells (Jan and Jan, 1993; Weintraub, 1993; Lee, 1997; Kintner, 2002; Parker, Seale and Rudnicki, 2003; Lee *et al.*, 2005; Lau *et al.*, 2018).

Classification of cell types

Historically, a large amount of effort has been put towards characterizing and categorizing the vast amount of different cell types generated during development. At present, however, no consensus of what comprises a cell type and how it should be curated is established, though many different viewpoints on this matter exist (Various Authors, 2017; Samantha A Morris, 2019; Xia and Yanai, 2019). Nevertheless, a useful framework to guide discussion on cell types and their identity has recently been put forward, consisting of three pillars: 1) Phenotype and function 2) Lineage and 3) State (**Figure 5**) (Samantha A Morris, 2019).

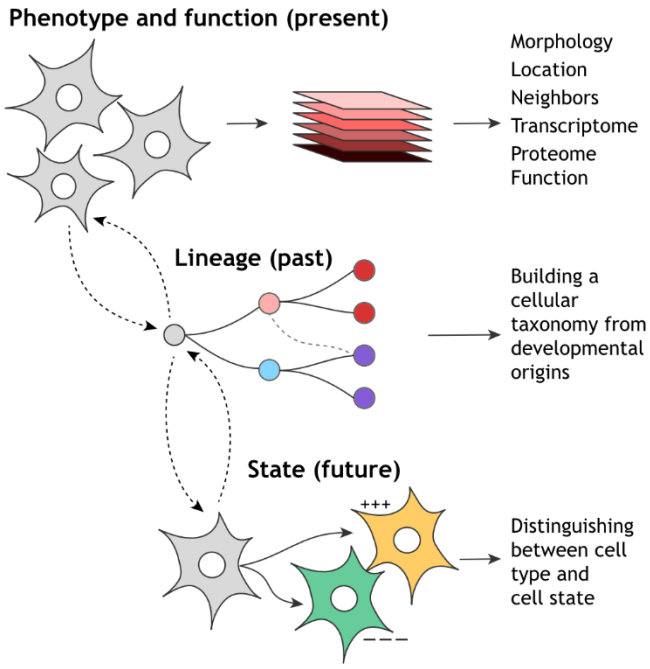


Figure 5: The three pillars of cell identity. Depicted are the three pillars of cell identity according to (Samantha A. Morris, 2019). Cell phenotype and function describe the set of measurable characteristics of a cell while lineage describes its developmental origin. Cell state on the other hand refers to the future of the cell and needs to be distinguished for cell type. Adapted from (Samantha A. Morris, 2019) with permission from Development.

Phenotype and function

Cellular phenotype and function refer to the diverse range of measurable characteristics that allow a systematic classification of cell types. Phenotype can be described by a wide variety of different techniques, ranging from marker-based imaging and flow-cytometry methods to more unbiased technologies such as the profiling of the transcriptome, epigenome and proteome. In particular, recent technological advances in the field of single cell genomics have resulted in an unprecedented resolution and scale at which the transcriptome and epigenome of cells can be profiled (Svensson, Vento-Tormo and Teichmann, 2018). Furthermore, one of the major shortcomings of these techniques, the loss of spatial information, has recently also been partially overcome with the advent of spatial transcriptomics (Ståhl *et al.*, 2016; Rodriques *et al.*, 2019; Vickovic *et al.*, 2019; Stickels *et al.*, 2020; Longo *et al.*, 2021; Marx, 2021). As a result, we currently have access to an exceptionally well-rounded toolbox to profile cellular phenotypes. Ultimately, however, the identity of a cell is best described by its function. In this regard, one of the major questions is to what extent molecular signatures derived by phenotypic methods correspond to functional cell types. Recently, several studies have started to address this question, by integrating the molecular characterization of cells with functional read outs (Gouwens *et al.*, 2020; Callaway *et al.*, 2021). These reports showed that there is a substantial

degree of correlation between molecular identities and cellular physiology, morphology and connectivity when applied to neuronal cells of the cerebral cortex (Gouwens *et al.*, 2020; Callaway *et al.*, 2021). This suggests that there is a strong molecular underpinning of cell identity, brought about by the developmental processes outlined above, which can be captured by single cell genomic methods at adult stages. In combination with the large scale at which these techniques can be performed, they represent an exceptionally powerful approach to establish a general framework for cell identity. They are, however, not perfect as the connectivity and activity of neurons was found to further refine their identity, an aspect poorly captured by single cell techniques. Taken together, these studies illustrate the necessity of collecting data at a large scale as well as integrating several modalities in order to reliably and thoroughly characterize and catalogue the fine-grained differences between (neuronal) cell types.

Lineage

In order to fully understand a cell's identity, it is important to understand its relation to other cells. For previously undescribed cell types, this might help infer some of the possible functions, while for more well-known cell types it is a powerful yet simple way to put the cell in a more general context. For this purpose, lineage tracing, the identification and tracking of cellular offspring, has been widely used (Kretzschmar and Watt, 2012). Although historically heavily relying on imaging-based methods, the advent of sequencing techniques has also made its mark here. Mainly, several different barcoding approaches have provided new ways in which cellular origin can be delineated (Raj *et al.*, 2018; Wagner *et al.*, 2018; McKenna and Gagnon, 2019; Bandler *et al.*, 2021). Amongst other things, these studies have provided substantial evidence for convergence, i.e., clonally distinct progenitors giving rise to similar cell types (Cao *et al.*, 2019; Bandler *et al.*, 2021), as well as divergence, i.e., clonally related cell diverging towards very distinct cell types (Sulston *et al.*, 1983; Bandler *et al.*, 2021). Hence, lineage tracing studies provide invaluable insight into how cell type diversity is established. Furthermore, the combination of lineage tracing with phenotypic data is expected to become an increasingly powerful approach due to the advent of spatial transcriptomic methods that can supplement lineage maps with spatial information (Samantha A Morris, 2019)

State

Finally, it is important to be able to distinguish cell identity from cell state. Cell identity can be thought of as a cell intrinsic property and is generally stably maintained by transcription factors (Holmberg and Perlmann, 2012; Samantha A. Morris, 2019). In neurons, for example, the same

transcription factors that initiate terminal differentiation, so-called terminal selectors, are also of vital importance for the maintenance of identity (Deneris and Hobert, 2014; Hobert and Kratsios, 2019). Cell state, on the other hand, is more fluid and cells can display a spectrum of different phenotypes (Samantha A. Morris, 2019). A prime example of this are T-cells, which can adopt a range of different activation states while maintaining their intrinsic T-cell identity (Zemmour *et al.*, 2018). This raises the question how one can distinguish between cell state and cell identity. Although this might be feasible for known cell types, it becomes considerably more difficult for unknown cell types. A further complicating matter for using unbiased transcriptomic to resolve this, is that not all variations in transcripts are meaningful, as transcription is often stochastic leading to heterogeneity among cells that is not necessarily related to cell state (Femino *et al.*, 1998; Costelloe *et al.*, 1999; Hume, 2000). To address both issues, the latest developments in the spatial profiling of transcriptomes, as well as approaches integrating different modalities (i.e., RNA, protein, and chromatin openness) will play an important role by offering increased resolution (Ståhl *et al.*, 2016; Rodriques *et al.*, 2019; Vickovic *et al.*, 2019; Zhu, Preissl and Ren, 2020). However, it is essential that such data is collected on a large scale to profile the entire range of variation in transcriptomic states in order to distinguish different cell states as well as detect potentially small populations of new cell types.

Plasticity of cell identity

Despite the ongoing efforts to document all cell types and their possible cell states, it was generally accepted that, once attained, cell identity was irreversible (Weismann, 1893). This idea, initially proposed by Weismann, was also incorporated by Conrad Waddington in his epigenetic landscape (**Figure 1**) (Weismann, 1893; Waddington, 1957).

Somatic Nuclear Transfer

Initial evidence to support the concept that differentiation is irreversible was provided by Robert Briggs and Thomas King (Briggs and King, 1952). In 1952, they established a technique known as nuclear transfer, where the nucleus of one cell is removed and replaced with the nucleus of another cell. Using this technique in their model system, the northern leopard frog (*Rana pipiens)*, they demonstrated that normal tadpoles could be generated when transferring the nucleus of a blastomere (totipotent cell of the early embryo) to an enucleated egg (Briggs and King, 1952). However, when the same experiment was performed with a nucleus derived from cells of the gastrula stage, which is more lineage restricted, this was no longer possible (King and Briggs, 1955). Hence, this suggested that the potential of a nucleus to support

development decreases with differentiation, i.e., there is an '*intrinsic restriction in potentiality for differentiation'* from the gastrula stage onwards (King and Briggs, 1955). However, shortly after these experiments were performed, John Gurdon and colleagues challenged this concept. Using a different species of frog, *Xenoypus leavis*, they demonstrated that the nucleus from a tadpole intestinal cell could drive development into fertile mature animals when it was implanted into enucleated oocytes (Gurdon, Elsdale and Fischberg, 1958; Gurdon and Uehlinger, 1966). This seminal work thereby demonstrated that differentiation does not result in the loss of genomic information and somatic nuclei maintain the capacity to support the development of a functional organism. Despite this proof of concept in amphibians, the first successful attempt in mammals was only achieved three decades later and its low efficiency prompted the question whether perhaps a population of contaminating stem cells could be responsible (Campbell *et al.*, 1996; Wakayama *et al.*, 1998, 1999; Gurdon and Byrne, 2003). To investigate this, adult B- and T-cells were used as donor cells since they undergo substantial DNA rearrangements that can be detected in clones. Although mice could not be directly generated when transferring B- and T-cell nuclei into enucleated oocytes, generation of embryonic stem cells (ESCs) from nuclear transfer derived blastocysts and subsequent injection into tetraploid embryos gave rise to fully functional animals in which the genomic rearrangements could be detected (Hochedlinger and Jaenisch, 2002). Hence, somatic nuclear transfer experiments were among the first to definitively demonstrate that differentiated cells maintain the capacity to revert to a pluripotent state.

Cell fusion

Another important takeaway from somatic nuclear transfer was that certain factors present within the cellular environment of pluripotent cells seemed to be capable of driving this dedifferentiation. Leveraging this, several groups started to perform cell fusion experiments, generating a hybrid cell type with two intact nuclei, referred to as a heterokaryons (Blau and Blakely, 1999). These studies were mainly focused on fusing pluripotent mouse embryonic stem cells (ESCs) with human somatic cells as a means of generating human pluripotent cells (Pralong, Trounson and Verma, 2006; Malinowski and Fisher, 2016; Imai *et al.*, 2020). Indeed, such experiments showed that transcription factors involved in the pluripotent gene regulatory network, such as Oct4 and Sox2, were induced in the resulting cell type (Pralong, Trounson and Verma, 2006; Malinowski and Fisher, 2016; Imai *et al.*, 2020). This was considered an important breakthrough and provided a more practical approach for obtaining pluripotent human cells, circumventing the logistical and ethical concerns related to obtaining human oocytes for somatic nuclear transfer. Nevertheless, interest quickly faded when the first paper demonstrating the generation of human induced pluripotent stem cells (see below) was

published. Recently, however, interest has sparked once again as hematopoietic stem cells derived from the bone marrow were found to convert their fate by fusion with somatic cells in a wide array of different tissues, including the heart (Alvarez-Dolado *et al.*, 2003), CNS (Weimann *et al.*, 2003; Johansson *et al.*, 2008), and liver (Alvarez-Dolado *et al.*, 2003; Vassilopoulos, Wang and Russell, 2003; Wang *et al.*, 2003). Moreover, it has become increasingly clear that such fusion events can contribute to the repair of damaged tissue (Spees *et al.*, 2003; Rizvi *et al.*, 2006; Davies *et al.*, 2009; Powell *et al.*, 2011). However, as such cell fusion events are extremely rare, they are unlikely to have a meaningful impact and thus call for further research to determine whether the process can be stimulated to promote functional recovery of tissue damage.

Reprogramming

The fact that both somatic nuclear transfer and cell fusion were able to instruct pluripotency in somatic cells further strengthened the idea that specific factors that drive dedifferentiation must be present in pluripotent cells. Hence, this initiated the search for such factors and led to one of the most important biological discoveries of the 21st century. In 2006, Kazutoshi Takahashi and Shinya Yamanaka reported that the expression of 24 ESC-derived genes could induce pluripotency when ectopically expressed in adult mouse somatic cells (Takahashi and Yamanaka, 2006). However, their most striking finding was that they identified a much smaller panel of four factors, Oct4, Sox2, Klf4 and c-Myc (OKSM), that was sufficient to induce pluripotency (Takahashi and Yamanaka, 2006). Hence, the entire cellular gene regulatory network of a differentiated somatic cell could be remodeled to a pluripotent state by just four factors (**Figure 6**). Although initially shown in mouse cells, the same factors were also sufficient to reprogram human adult somatic cells to pluripotency (Park *et al.*, 2007; Takahashi *et al.*, 2007; Yu *et al.*, 2007). Since then, numerous studies have helped shed light on the mechanisms of this conversion (Takahashi and Yamanaka, 2006; Maherali *et al.*, 2007; Mikkelsen *et al.*, 2008; Sridharan *et al.*, 2009; Spitz and Furlong, 2012; Heinz *et al.*, 2015; Zunder *et al.*, 2015; Chen *et al.*, 2016; Chronis *et al.*, 2017; Knaupp *et al.*, 2017; Li *et al.*, 2017; Nefzger *et al.*, 2017; Schwarz *et al.*, 2018; Francesconi *et al.*, 2019; Guo *et al.*, 2019; Schiebinger *et al.*, 2019; Tran *et al.*, 2019; Liu *et al.*, 2020; Xing *et al.*, 2020). In short, fibroblasts first undergo a transition towards a more epithelial state, from which only a small proportion of cells successfully reprogram to induced pluripotent stem cells (iPSCs), whereas the majority of cells do not proceed beyond this point or adopt alternative fates (Mikkelsen *et al.*, 2008; Li *et al.*, 2010; Samavarchi-Tehrani *et al.*, 2010; Hussein *et al.*, 2014; Parenti *et al.*, 2016; Schiebinger *et al.*, 2019; Liu *et al.*, 2020; Xing *et al.*, 2020). For cells that do reprogram, fibroblast identity is gradually repressed while the pluripotency program is activated following the stepwise activation of pluripotency

transcription factors and their downstream targets that establish the pluripotency gene regulatory network (Buganim *et al.*, 2012; Chronis *et al.*, 2017; Deng *et al.*, 2021).

Therapeutic potential of iPSCs

The discovery of iPSCs provided researchers with an indefinite source of human pluripotent cells and offered tremendous potential for disease modelling and personalized medicine. Indeed, since their discovery more than 15 years ago, iPSCs have contributed to significant progress in these areas (Rowe and Daley, 2019). For example, better understanding of disease mechanisms using patient derived iPSCs has resulted in the repurposing of existing drugs for several diseases including Alzheimer's disease, amyotrophic lateral sclerosis and spinal muscular atrophy (Bright et al., 2015, McNeish et al., 2015, Naryshkin et al., 2014, Wainger et al., 2014). Furthermore, the first cell replacement therapies using human iPSC derived dopaminergic cells for the treatment of Parkinson's disease have also recently begun, with additional efforts currently in the pipeline (Barker *et al.*, 2017; Cyranoski, 2018; Schweitzer *et al.*, 2020). However, many diseases are not cell autonomous in nature and are instead contributed to by multiple cell types. Acknowledging this, the field has shifted focus towards the generation and study of self-organizing multicellular 3D tissue models that mimic organ function, collectively referred to as organoids (Lancaster *et al.*, 2013; Quadrato *et al.*, 2017; de Souza, 2018; Velasco *et al.*, 2019; Kim, Koo and Knoblich, 2020). Remarkably, several labs have recently even reported the fusion and functional integration between different organoid structures to give rise to even more complex 3D structures called assembloids (Bagley *et al.*, 2017; Birey *et al.*, 2017; Xiang *et al.*, 2017; Cakir *et al.*, 2019; Faustino Martins *et al.*, 2020; Vogt, 2021). For example, organoids with different regional brain identities have been fused together to form neural assembloids that facilitate the study of how connectivity is established during neuronal development (Bagley *et al.*, 2017; Birey *et al.*, 2017; Xiang *et al.*, 2017). In addition, similar efforts on different organ systems have also been published recently (Kim *et al.*, 2020; Rawlings *et al.*, 2021). Overall, the continuous development of these techniques is expected to give rise to further complexity, increasing our ability to recapitulate human biology. However, despite these encouraging developments, several challenges surrounding the use of iPSCs for disease modelling and cell replacement strategies persist. One shortcoming, for instance, is the fact that iPSC generation resets the epigenetic state of the cell to a juvenile stage (Lapasset *et al.*, 2011; Horvath, 2013; Miller *et al.*, 2013). This is particularly relevant for the modelling of neurodegenerative diseases, as the inability of neurons to divide makes them especially susceptible to aging and is thought to underlie the late onset of many neurodegenerative diseases (Hou *et al.*, 2019; Azam *et al.*, 2021). iPSC derived neurons, however, fail to capture this risk factor due to their immature phenotype (Mertens *et al.*, 2015; Tang *et al.*, 2017). Genetic

heterogeneity between different iPSC lines poses another challenge. Several studies have shown that there is substantial variation in the phenotype of differentiated cells due to background genetic variation (DeBoever, 2017; Kilpinen *et al.*, 2017; Pashos *et al.*, 2017; Warren, 2017). Such variations have a confounding effect on disease modelling and call for the use of isogenic controls, where the genetic abnormalities are corrected to create an otherwise identical genomic background. Though this might be possible for monogenic disorders, it is difficult to achieve for more complex polygenic diseases or diseases where the exact genetic component is currently unknown. Fortunately, alternative reprogramming methods that circumvent some of these issues have seen a similar degree of progress alongside that of iPSC-based reprogramming.

Transdifferentiation

In particular, transdifferentiation, a method first described in the 1980s, has seen a tremendous amount of progress and offers several major advantages over iPSC reprogramming (Davis, Weintraub and Lassar, 1987). Like iPSCs, transdifferentiation (also known and hereafter referred to as direct reprogramming) relies on the forced overexpression of transcription factors to drive conversion. However, in contrast to iPSCs, it directly converts one cell type into the other without passing through a pluripotent stage (**Figure 6**)(Davis, Weintraub and Lassar, 1987; Heins *et al.*, 2002; Heinrich *et al.*, 2010; Vierbuchen *et al.*, 2010; di Tullio and Graf, 2012; Fishman *et al.*, 2015; Treutlein *et al.*, 2016). First of all, this substantially decreases the risk of tumor formation. Second, as there is no need to generate and expand pluripotent progenitors before (trans)differentiation, it is generally much faster as well. As such, studies have been able to include a far greater number of donors and thus have been able to better account for genetic heterogeneity. Third, direct reprogramming maintains the aging marks present in the starter cells (Mertens *et al.*, 2015; Huh *et al.*, 2016; Victor *et al.*, 2018), a major benefit for the modelling of diseases with a late onset. Finally, the direct nature of the approach also offers a unique opportunity for *in vivo* repair strategies, as cells can be directly converted into the desired cell type within the tissue of interest, which impossible to achieve with iPSCs.

Direct conversion was first observed by chance when a potential chemotherapeutic agent was found to convert fibroblast-like cells into myoblasts (Constantinides, Jones and Gevers, 1977; Taylor and Jones, 1979). Further investigation through the screening of myogenic cDNA clones led to the discovery of the MyoD1 gene, which was found to be responsible for driving this conversion (Taylor and Jones, 1979; Davis, Weintraub and Lassar, 1987). However, technical difficulties as well as conceptual concerns initially slowed down the techniques' further development. From the start of the 21st century however, several successive seminal findings

were made. First, it was shown that ectoderm derived glia could be converted into functional neurons (Heins *et al.*, 2002). Next, two successive studies demonstrated that pre-T cells and B-cells could be converted into macrophages (Xie *et al.*, 2004; Laiosa *et al.*, 2006). Shortly thereafter, pancreatic cells were also found to be reprogrammable into insulin secreting beta cells (Zhou *et al.*, 2008). Notably, these initial experiments were limited to conversions between cell types with a similar germ layer identity, as this was conceptually considered most feasible (Hochedlinger and Plath, 2009). However, not much later, cross germ layer conversions were also demonstrated when skin derived fibroblasts were shown to be converted into neurons (Vierbuchen *et al.*, 2010) as well as hepatocytes (Huang *et al.*, 2011; Sekiya and Suzuki, 2011). Since then, the number of protocols for conversion within and across germ layers has been steadily increasing and many protocols are now available for a plethora of different starter cells derived from both mice and humans (reviewed by H. Wang et al., 2021).

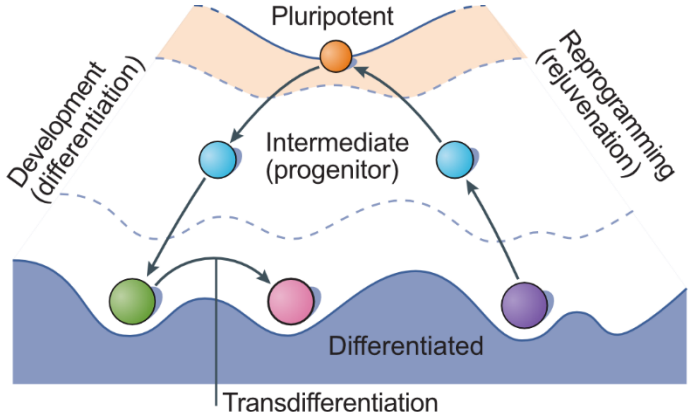


Figure 6: Waddington's landscape revisited. Adaptation of the original Waddington's landscape to incorporate the recent findings on reprogramming and transdifferentiation. Adapted from Takahashi and Yamanaka, 2006 with permission from Springer Nature (License No. 5235390867893).

Reprogramming factors

To identify transcription factors that have the potential to act as reprogramming factors, researchers have initially heavily relied on knowledge from embryonic development to narrow down possible candidates. For example, after the initial discovery of MyoD1 and its potential in both myogenesis and reprogramming, its counterpart Myf5 was discovered to have similar capabilities (Davis, Weintraub and Lassar, 1987; Dimicoli-Salazar *et al.*, 2011; Watanabe *et al.*, 2011; Chakraborty *et al.*, 2014; Dall'Agnese *et al.*, 2019). Furthermore, a comparison between the myogenic trajectories during embryogenesis and myogenic reprogramming revealed both processes are highly similar, although many cells diverge towards unproductive outcomes during reprogramming (Cacchiarelli *et al.*, 2018). Knowledge from development was applied in

a similar way for the neurogenic lineage. Following the observation that the neurogenic factor Pax6 could convert astrocytes to neuronal cells, its known downstream target Neurog2 as well as its counterpart Ascl1 were also investigated and found to convert astrocytes into neurons even better (Heins *et al.*, 2002; Berninger *et al.*, 2007; Heinrich *et al.*, 2010). In line with their developmental roles, these factors instructed very different programs in cerebral cortex derived astrocytes and ultimately established GABAergic and glutamatergic identities respectively (Berninger *et al.*, 2007; Heinrich *et al.*, 2010; Masserdotti *et al.*, 2015). However, in contrast to MyoD1, both factors were also capable of converting more distantly related cells into neurons, namely fibroblasts and hepatocytes (Davis, Weintraub and Lassar, 1987; Weintraub *et al.*, 1989; Vierbuchen *et al.*, 2010; Marro *et al.*, 2011; Chanda *et al.*, 2014; Smith *et al.*, 2016). Between the two factors, Ascl1 was found to be most potent in this regard, as it could convert fibroblast into neurons on its own (Vierbuchen *et al.*, 2010; Marro *et al.*, 2011; Chanda *et al.*, 2014). Neurog2, on the other hand, was unable to convert fibroblast into neurons directly without the addition of neurogenic small molecules (Smith *et al.*, 2016). Interestingly, although both factors instruct opposite neuronal identities during development and reprogramming, their joint expression resulted in increased reprogramming efficiencies in fibroblasts (Ladewig *et al.*, 2012; Mertens *et al.*, 2015).

Another example of how studying development has served as the basis for the identification of reprogramming transcription factors is the identification of members from the FoxA and Hnf family by screening factors involved in liver development (Rombaut *et al.*, 2021). FoxA2, together with the central regulator of liver differentiation Hnf4α (Watt, Garrison and Duncan, 2003) was the first combination found to be capable of converting mouse fibroblasts into induced hepatocytes (Sekiya and Suzuki, 2011). Similarly, FoxA3 together with Gata4 and Hnf1α could achieve the same in a non-proliferative background (Huang *et al.*, 2011). Strikingly, Hnf1α on its own was also shown to be capable of achieving hepatic conversion, though it was aided by a considerable number of small molecules (Lim *et al.*, 2016). Functionally, however, the generated cells in all protocols were found to substantially differ from primary hepatocytes and were unable to rescue liver function, as they were more similar to endoderm progenitors and showed the plasticity to be engrafted into other endoderm derivatives (Du *et al.*, 2014; Morris *et al.*, 2014; Pournasr, Asghari-Vostikolaee and Baharvand, 2015; Lim *et al.*, 2016; Orge *et al.*, 2020). Nevertheless, this revealed that, much like during development, FoxA factors are key to instructing a general competence for differentiation into the hepatic lineage, while relying on additional factors to drive proper maturation (Gualdi *et al.*, 1996). The question remained, however, what bestows this shared capacity to drive fate transitions upon the different transcription factors discussed here.

Pioneer factors

The answer to the abovementioned question was found to partially lie in their unique capacity to interact with DNA. To appreciate this, it is important to understand how structural differences in chromatin, the combined complex of DNA and proteins, can regulate gene expression. To fit within the nucleus, DNA is not freely distributed, but is instead wrapped around histones to form arrays of so-called nucleosomes (Kornberg and Lorch, 1999). DNA facing these histones is inherently inaccessible due to steric hindrance and provides a first way of regulating gene expression (Luger *et al.*, 1997). Furthermore, the array of nucleosomes can either condense, resulting in tightly packed chromatin that is inaccessible for transcription, or adopt a looser conformation, reminiscent of beads on a string (Schalch *et al.*, 2005). The latter is referred to as euchromatin and is generally considered as open, i.e., accessible for transcription (Huisinga Brent Brower-Toland Sarah C R Elgin, 2006). Hence, regulation of chromatin condensation represents a powerful mechanism via which gene expression states can be stabilized. In the context of cell identity, gene expression of alternative cell fates is often silenced through chromatin condensation. However, among the above-mentioned transcription factors, Ascl1, Neurog2, MyoD1, FoxA2, Sox2 and Oct4, belong to a group of pioneer transcription factors with the unique ability to engage their target sites in closed or unmarked chromatin (Iwafuchi-Doi and Zaret, 2014; Morris, 2016; Zaret and Mango, 2016) (**Figure 7A**). Following their binding, they can increase the accessibility of chromatin to other transcription factors as well as various other factors that influence chromatin condensation status (**Figure 7B**), generally resulting in chromatin opening and cooperative binding to drive transcription (**Figure 7C**)(Zaret, 2020). However, despite the fact that they are grouped under the same name, pioneer factors display very different mechanisms to achieve their outcome, as structural differences between factors influence the nature of their interaction with the DNA (Mayran and Drouin, 2018). As a result, it has been suggested that some factors achieve similar binding of their targets across cell types, while others show impaired binding when target sites are enriched with repressive chromatin marks (Soufi, Donahue and Zaret, 2012; Wapinski *et al.*, 2017; Q. Y. Lee *et al.*, 2020). For this reason, the term 'on-target' pioneer factor was conceived to refer to the former class of pioneer factors. However, strong evidence to support this claim is currently lacking and a systematic comparison of the binding behavior of pioneer factors under different cellular circumstances has not been performed so far.

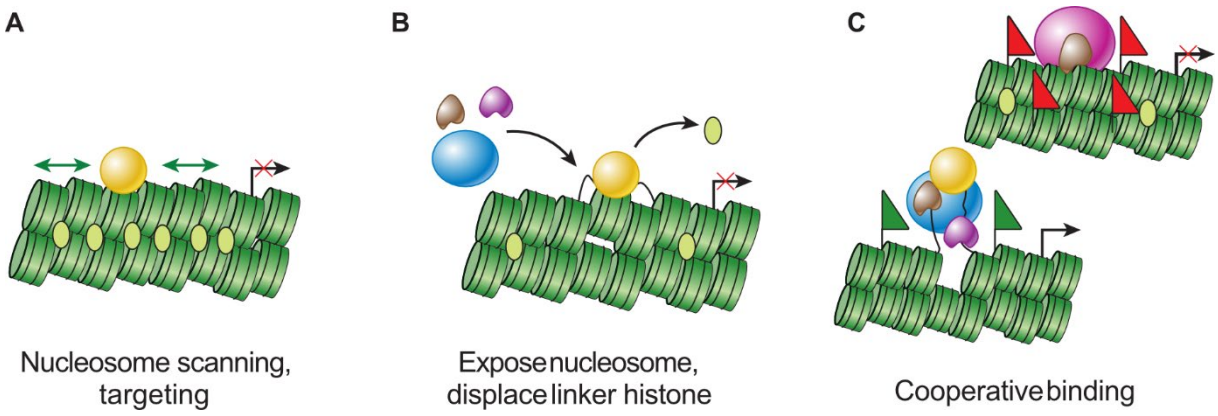


Figure 7: Pioneer factor function. Schematic overview of how pioneer factors scan nucleosomes (a) displace linker histones to expose the DNA (b) and facilitate cooperative binding to allow transcription (c). Adapted from Zaret, 2020 with permission from Annual Review (License No. 1181464-1).

Starter cells in reprogramming

The pioneer activity of these factors thus provided a mechanistic explanation for the widespread applicability and potency of the same factors across different starter cells. Indeed, the current diversity of cell types that can be generated through reprogramming and the large amount of different starter cells that can be used to achieve this, suggests that any cell might be converted into any other cell type given the right combination of factors (Xu, Du and Deng, 2015; Morris, 2016; Wang *et al.*, 2021). Nevertheless, two main cell types have been of particular value to the field of reprogramming: skin derived fibroblasts and astrocytes (Xu, Du and Deng, 2015).

Dermal fibroblasts

Fibroblasts are mesoderm derived mesenchymal cells that secrete extra cellular matrix components and collagens to support organ function (Plikus *et al.*, 2021). Although they can be found in several organs, such as the skin, heart and lungs, skin derived fibroblasts have been the preferred starter cell for many reprogramming paradigms due to the relative ease with which they can be acquired and cultured from both mice and humans (Durkin *et al.*, 2013; Kisiel and Klar, 2019). This makes them particularly useful for disease modelling, as patient derived cells can be easily acquired through skin biopsies. Furthermore, they have been proven to be a versatile starter cell that can be reprogrammed into a large variety of different cell types (Morris, 2016; Wang *et al.*, 2021). Exceptionally valuable, however, has been the advent of direct fibroblast to induced neuron (iN) reprogramming in human cells (Vierbuchen *et al.*, 2010; Pang *et al.*, 2011; U. Pfisterer *et al.*, 2011; Treutlein *et al.*, 2016). First of all, because *in vivo*

mouse models are limited in their ability to recapitulate the intricate phenotypes of human neurological diseases (Ransohoff, 2018). Second of all, because the availability of human post-mortem brain tissue is severely limited and does not allow for any functional studies to be performed. In addition, donor derived iNs were shown to maintain hallmarks of transcriptomic, epigenetic and metabolic aging, in contrast to their iPSC derived counterparts (Mertens *et al.*, 2015; Huh *et al.*, 2016; Tang *et al.*, 2017; Kim *et al.*, 2018; Victor *et al.*, 2018). As such, they were deemed to be the most suitable cell type for the modelling of human neurodegenerative disease, for which age is the main risk factor. Indeed, modelling several human neurological diseases using iNs recapitulated several important aspects of the disease phenotype, whereas their rejuvenated iPSC counterparts did not (Tang *et al.*, 2017; Victor *et al.*, 2018). As a result, the development of iN based models has substantially increased our understanding of the mechanisms underlying several neurological diseases (Liu *et al.*, 2014; Liu, Zang and Zhang, 2016; Kim *et al.*, 2017; Zhang *et al.*, 2017; Pircs *et al.*, 2018, 2021; J. E. Lee *et al.*, 2020; Brattås *et al.*, 2021). More importantly, it is expected that the use of such models will substantially aid the development of therapeutic strategies in the future. Taken together, the relative ease with which they can be isolated as well as their tremendous versatility in reprogramming have made fibroblasts an invaluable cell type for expanding our knowledge on (human) cell function in health and disease.

Astrocytes

In contrast to fibroblasts, interest in astrocytes has mostly been driven by their potential for tissue repair. One of the reasons for this is their widespread localization and abundance within the CNS, making them an ideal *in loco* source to replace lost neurons. Another important consideration is that astrocytes and neurons share a common ancestor, possibly facilitating neuronal conversion. This common ancestor, the radial glia cells, derive from neuroepithelial cells lining the neural tube and are transiently found in most brain regions (Kriegstein and Götz, 2003; Anthony *et al.*, 2004). During the early stages of neuronal development, radial glia predominantly give rise to neuronal progeny (Miller and Gauthier, 2007; Rowitch and Kriegstein, 2010). However, as development progresses, a subset of radial glia undergoes an extensive genetic switch that marks the advent of gliogenesis, a process that extends into postnatal stages and sequentially gives rise to astrocytes and oligodendrocytes (Miller and Gauthier, 2007; Molofsky and Deneen, 2015). Once formed, astrocytes further divide locally to expand their numbers (Ge *et al.*, 2012). Consequently, astrocytes are an abundant cell type found scattered throughout the entire CNS where their highly ramified morphologies allow them to establish contact with other glial cells, blood vessels and synapses (Bushong *et al.*, 2002; Halassa *et al.*, 2007). Of all the contacts they establish, their contact with synapses is

considered exceptionally important, as astrocytes not only provide neurotransmitter, ion and general volume homeostasis but also isolate individual synapses from each other to ensure the fidelity of synaptic transmission (Verkhratsky and Nedergaard, 2014, 2018). In addition, they provide neurons with energy metabolites such as glucose and lactate as well as trophic support (Gomes, Paulin and Neto, 1999; Qu *et al.*, 2000; Rouach *et al.*, 2008; Figley and Stroman, 2011; Dezonne *et al.*, 2013; Sotelo-Hitschfeld *et al.*, 2015). Furthermore, astrocytes can also actively modulate neuronal activity through the secretion of gliotransmitters into the synaptic cleft. As such, astrocytes, are considered so inherent to the function of synapses that, together with the pre- and post-synaptic neuronal membrane, they are often referred to as a single functional unit called the tripartite synapse or synaptic cradle (Araque *et al.*, 1999; Verkhratsky and Nedergaard, 2014). Astrocytes, however, do not only facilitate neuronal communication, but also directly communicate with one another over long distances through the propagation of calcium waves via gap-junctions (Kanemaru et al 2014, De Bock et al 2014, Scemes and Giaume 2006). Although the exact physiological roles of these waves are not fully understood, there are indications that they can influence both neuronal physiology and behavior (Chai et al 2017, Yu et al 2018). Furthermore, astrocytes to a large extent also control general homeostasis in the brain by regulating ion, water and pH levels while their contact with the vasculature strongly contributes to the maintenance of the blood brain barrier (Rothstein *et al.*, 1994, 1996; Volterra and Meldolesi, 2005; Siqueira *et al.*, 2018). In summary, astrocytes fulfill several different functions, highly dependent on the cells they are in contact with, making them an essential cell for proper brain functioning (Barres, 2008; Allen and Barres, 2009; Eroglu and Barres, 2010; Khakh and Sofroniew, 2015; Allen and Eroglu, 2017; Allen and Lyons, 2018; Khakh, 2019; Khakh and Deneen, 2019; Santello, Toni and Volterra, 2019).

However, in addition to their roles in the healthy brain, astrocytes also play a central role in the response to CNS injury. This behavior of astrocytes was first reported in the late 19th century, when several neuroanatomists first noted that astrocytes undergo pronounced structural changes in response to damage within the CNS. They referred to this phenomenon as astrogliosis, a term that is currently still widely used (Eddleston and Mucke, 1993; Sofroniew, 2009, 2015; Sofroniew and Vinters, 2010; Clarke *et al.*, 2021). However, usage of a single term falsely suggests that astrogliosis is a single phenotype that is the same under all conditions. Instead, it comprises a wide spectrum of changes that can range from mild reversible changes in gene expression and cellular hypertrophy to cell proliferation and permanent changes in tissue arrangement (**Figure 8**) (Sofroniew, 2009). Another topic of common debate is whether astrogliosis is beneficial or detrimental to functional recovery. At its core, astrogliosis is a beneficial process that is aimed at promoting wound healing and protecting the tissue from

further damage (Sofroniew, 2009, 2015). However, it can be harmful under certain conditions. One example is the formation of a glial scar, which results from severe astrogliosis and not only has a strong negative effect on the outgrowth of new axons but also causes permanent tissue rearrangements that are generally detrimental to recovery (**Figure 8**) (Silver and Miller, 2004). For this reason, it has been suggested that inhibiting astrogliosis might be beneficial to recovery and could be considered as a potential target for therapeutic strategies. However, this notion relies on the assumption that astrogliosis and scar formation are detrimental by definition, something that has been contested by many studies showing that lack of scar formation results in even poorer outcomes (Nawashiro *et al.*, 1998; Bush *et al.*, 1999; Faulkner *et al.*, 2004; Myer *et al.*, 2006; Drögemüller *et al.*, 2008; Herrmann *et al.*, 2008; Li *et al.*, 2008; Voskuhl *et al.*, 2009; Haroon *et al.*, 2011; Macauley, Pekny and Sands, 2011; Wanner *et al.*, 2013). Therefore, any therapeutic strategies should be aimed at attenuating only some aspects of astrogliosis, such as excessive scar formation, while leaving most other aspects untouched. In this regard, *in vivo* direct neuronal reprogramming of (reactive) astrocytes might prove to be one piece of the puzzle, as it could deplete astrocytes at the site of injury, possibly reducing the scale of glial scar formation, while at the same time promoting functional recovery through the generation of new neurons.

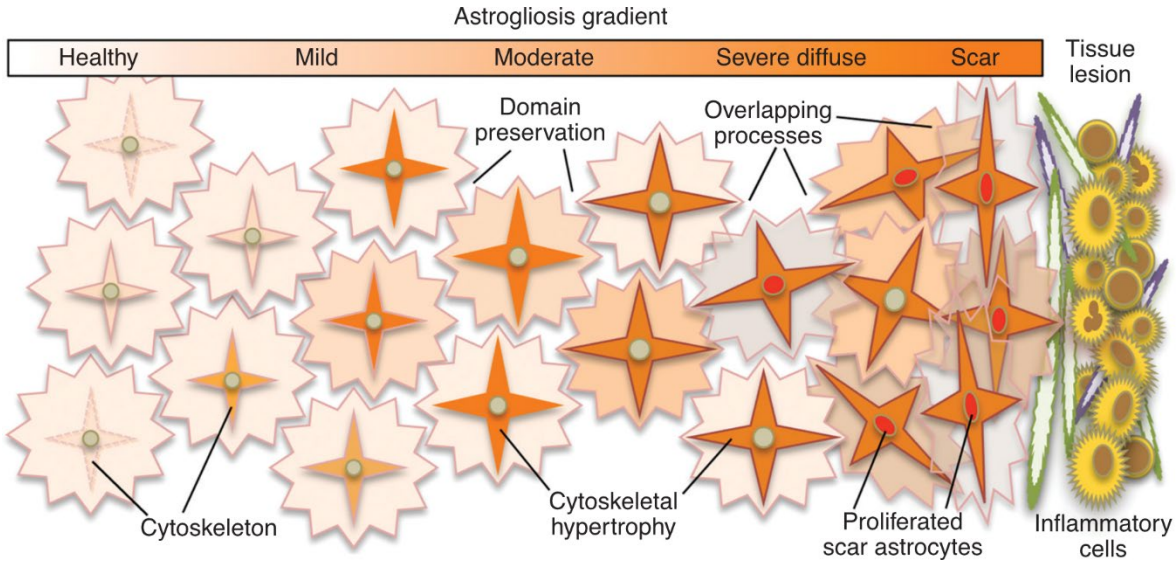


Figure 8: Astrogliosis. Schematic depiction of the astrogliosis spectrum from mild (left) to severe (right) and glial scar formation (far right). Adapted from (Sofroniew, 2015) with permission from Cold Spring Harbor Laboratory Press.

Astrocyte heterogeneity

As mentioned above, patterning of embryonic progenitors is a common mechanism via which cellular diversity is generated in different organs, including the CNS (**Figure 4**)(Jessell, 2000; Cohen, Briscoe and Blassberg, 2013; Bier and de Robertis, 2015). In addition, differentiation is often further refined by the interactions with neighboring cells to ultimately shape the final cell identity. Although these conserved principles have been well described for neurons (Jessell, 2000), whether or not they play a role in the diversification of astrocytes populations has historically received far less attention. Recently, however, research on astrocyte heterogeneity has gained momentum and an increasing number of studies have started to shed more light on this matter. For example, mutually exclusive expression of patterning transcription factors in progenitor populations was found to give rise to molecularly distinct subpopulations of astrocytes (Hochstim *et al.*, 2008; Tsai *et al.*, 2012). In addition, neuron derived signaling cues were demonstrated to also regulate astrocyte form and function (Stogsdill et al 2017, Farmer et al 2016). Given the regionalization of the brain and the different neuronal subtypes that exist within these regions, it was postulated that astrocytes may therefore display regional specialization. Indeed, astrocytes from different regions were shown to display extensive molecular heterogeneity and were found to preferentially support neuron function when derived from the same brain region (Morel, Ming Sum R Chiang, *et al.*, 2017). Furthermore, studies revealing different membrane properties, calcium signaling dynamics, ion balance and morphology as well as differences in the ability to connect to synapses and support neuronal maturation further highlighted the diversity in astrocyte function between regions (Chai *et al.*, 2017; Morel, Ming Sum R. Chiang, *et al.*, 2017; Lanjakornsiripan *et al.*, 2018). Interestingly, these regional differences seem to be rather stable, as they were also found in aging astrocytes (Boisvert *et al.*, 2018; Clarke *et al.*, 2018). In addition, recent studies have begun to dissect the heterogeneity of astrocytes within brain regions. One pioneer study isolated astrocytes from different brain regions and identified not only distinct molecular signatures in astrocytes between the regions but also identified subpopulations of astrocytes that were present in all regions (John Lin *et al.*, 2017). In other words, they found evidence for both inter- and intra-regional differences between astrocytes. More recent work has confirmed these inter-regional and even layer-specific morphological and molecular differences between astrocytes (Lanjakornsiripan *et al.*, 2018; Batiuk *et al.*, 2020; Bayraktar *et al.*, 2020; Ohlig *et al.*, 2021). Taken together, this reinforces the idea that astrocytes are prime candidates for neuronal replacement strategies as they might generate regionally appropriate neuronal subtypes that facilitate integration into existing circuits and thereby promote functional recovery.

Astrocyte to neuron conversion

Towards realizing the potential of astrocytes to replace lost neurons, astrocyte to neuron conversion first needed to be established. To this end, the first demonstration of astrocyte to neuron conversion was the neuronal conversion of postnatal cortical astrocytes upon the re-expression of the transcription factor Pax6, one of the key regulators of neurogenesis (Bishop, Goudreau and O'Leary, 2000; Heins *et al.*, 2002; Sansom *et al.*, 2009). As already briefly mentioned above, its downstream target Neurog2 was also found to be capable of achieving this and, strikingly, was substantially more efficient in doing so (Berninger *et al.*, 2007; Heinrich *et al.*, 2010). Similarly, its ventral forebrain counterpart Ascl1 was also found to convert astrocytes into neurons, generating cells with neuronal morphology and capable of firing action potentials (Berninger *et al.*, 2007; Heinrich *et al.*, 2010). These studies were the first to demonstrate the extraordinary ability of these factors to impose the neural fate in somatic cells and thereby paved the way for their application in different starter cells (see above). More important for the field of regenerative medicine, however, was that Neurog2 and Ascl1 generated glutamatergic and GABAergic neurons respectively, in accordance with the neuronal subtypes they specify in the cortex during development (Bertrand, Castro and Guillemot, 2002; Berninger *et al.*, 2007; Heinrich *et al.*, 2010; Masserdotti *et al.*, 2015; Hu *et al.*, 2019). This raised the question whether both factors might instruct the same program irrespective of regional identity or whether their behavior would be shaped by the starter cell to generate region specific neuronal subtypes. To test this, several studies have compared reprogramming of astrocytes derived from different regions of the CNS and have provided evidence to support the generation of regionally correct neuronal subtypes (Chouchane *et al.*, 2017; Hu *et al.*, 2019; Herrero-Navarro *et al.*, 2021; Rao *et al.*, 2021). For example, the regional identity of generated neurons could be manipulated through altering the 'regional-code' of the astrocytes (Herrero-Navarro *et al.*, 2021). Moreover, the maintenance of regional identity upon conversion seems to hold even when applied directly in tissue, as the conversion of astrocytes into region and even layer specific astrocytes *in vivo* has recently been reported (Mattugini *et al.*, 2019). However, although promising, the mechanisms via which regional astrocyte identity shapes the actions of reprogramming factors in regional astrocytes remains unknown. Nevertheless, the field of direct neuronal reprogramming has made tremendous progress over the past decades and is gradually getting closer to a stage where clinical trials might become feasible.

AIMS OF STUDY

The advent of reprogramming has been one of the major recent breakthroughs in the field of biology. However, the general principles underlying reprogramming, as well how starter cells influence it, are only partially understood. During my Ph.D. work, I focused on addressing some of the outstanding questions relating to this topic, namely: How is cell fate erased and re-established during the conversion process? How do different reprogramming factors achieve this in the very same starter cell? And conversely, how does the cell of origin influence direct neuronal reprogramming and the neuronal subtypes generated? I addressed these questions in two different but complementary projects.

In the first project, we set out to reveal the general overarching principles of reprogramming by starting from a single starter cell, primary cultures of mouse embryonic fibroblasts, using several different reprogramming factors. To this end, we first aimed to establish a novel single cell RNA-sequencing tool via which we could assess the early transcriptomic changes induced by different reprogramming factors under identical conditions. Upon establishment of this tool, our next aims were to determine: i) Whether different reprogramming factors utilize similar mechanisms for the removal of starter cell identity, ii) To what extent reprogramming factor levels influence the conversion process and iii) How cells resolve cell identity conflicts when more than one fate is imposed upon the cell at the same time.

In the second project we aimed to take the opposite approach, seeking to compare the actions of the same reprogramming factors in two different populations of starter cells. To this end, we first aimed to establish a highly enriched primary culture of astrocytes from the spinal cord. Once established, we intended to reprogram these cultures with Ascl1 and Neurog2 and characterize both the early events as well as the final outcome of reprogramming at the molecular level (electrophysiology and transcriptomics). Finally, we planned to compare this data with previously published data from the lab (Masserdotti *et al.*, 2015) where a similar analysis was performed on cortical-derived astrocytes. This would allow us to determine whether cortical and spinal cord astrocytes maintain their developmental patterning signature *in vitro* as well as if the same reprogramming factors instruct different programs when expressed in cortical and spinal cord astrocytes. As such, this would shed more light on the importance of the starter cell in shaping the reprogramming potential of transcription factors.

In summary, during my Ph.D. I aimed to investigate the mechanisms involved in fate conversion and thereby gain insights that could contribute to improving current direct reprogramming strategies.

RESULTS I

Probing cell identity hierarchies by fate titration and collision during direct reprogramming

Bob A. Hersbach*, David S. Fischer*, Giacomo Masserdotti, Deeksha, Karolina Mojžišová, Thomas Waltzhöni, Diego Rodriguez-Terrones, Matthias Heinig, Fabian J. Theis#, Magdalena Götz#, Stefan H. Stricker#

*#: These authors contributed equally

Author contributions

S.H.S conceived this project. The experimental approach was designed by S.H.S and **B.A.H** together with M.G., F.J.T., G.M. and D.S.F. **B.A.H.** performed all cloning and biological experiments. T.W., D.R. and M.H. performed genome alignment for sequencing data. D.S.F. and K.M performed all bioinformatic analysis. S.H.S, D.S.F and **B.A.H.** wrote the manuscript and all authors contributed to corrections and comments.

Permissions

Article under revision at EMBO Molecular Systems Biology as of June 28th, 2022.

Probing cell identity hierarchies by fate titration and collision during direct reprogramming

Bob A. Hersbach[1,2,3]*, David S. Fischer[4,6]*, Giacomo Masserdotti[1,2], Deeksha[1,2], Karolina Mojžišová[4], Thomas Waltzhöni[4,5], Diego Rodriguez-Terrones[4,8], Matthias Heinig[4,6], Fabian J. Theis[4,6,9#], Magdalena Götz[1,2,9#], Stefan H. Stricker[1,2#]

[1]Institute of Stem Cell Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

[2]Department of Physiological genomics, Biomedical Center Munich, Ludwig-Maximilians University, Planegg-Martinsried, Germany

[3]Graduate School of Systemic Neurosciences, Biocenter, Ludwig-Maximilians University, Planegg-Marrtinsried, Germany

[4]Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

[5]Core Facility Genomics, Helmholtz Zentrum München, Neuherberg, Germany

[6]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany.

[7]Department of Informatics, Technical University of Munich, Munich, Germany

[8]Current address: Research Institute of Molecular Pathology (IMP), Vienna BioCenter, Vienna, Austria

[9]German Excellence Cluster of Systems Neurology, Biomedical Center Munich, Planegg-Martinsried, Germany

*# These authors contributed equally to this work.

Correspondence to S.H.S. (Stefan.stricker@helmholtz-muenchen.de), F.J.T. (fabian.theis@helmholtz-muenchen.de) or M.G. (magdalena.goetz@helmholtz-muenchen.de)

Summary

Despite the therapeutic promise of direct reprogramming, a general concept on how cells resolve the cell identity conflicts underlying fate conversion is lacking. It is particularly unclear as to whether the erasure of original identities follows similar principles, whether reprogramming factor expression levels influence the outcome, and how reprogramming factors disturb each other's transcriptional program. To tackle these fundamental questions, we established a single-cell protocol for the simultaneous initiation and analysis of multiple cell fate conversion events based on combinatorial and traceable reprogramming factor expression (Collide-seq). Collide-seq revealed the lack of a common mechanism through which loss of fibroblast specific gene expression is triggered by different reprogramming factors. Moreover, the transcriptome of converting cells dramatically changed only when a critical level of each factor was attained, with higher or lower levels not contributing to major changes. By simultaneously inducing multiple competing reprogramming factors, Collide-seq revealed a deterministic system in which titration of fates against each other yields dominant or colliding fate pairs e.g., containing elements of both programs. By investigating one collision in detail, we find that reprogramming factors can disturb cell identity programs independent of their ability to bind their target genes. Taken together, Collide-seq has shed light on several fundamental principles of fate conversion that provide new insight and may aid further improving current reprogramming paradigms.

Introduction

Mammalian development, through a complex set of intrinsic and extrinsic signals, leads to a plethora of functionally distinct cell types characterized by gene regulatory networks that are stable under physiological conditions (**Figure 1A**) (Vickaryous and Hall, 2006; Enver *et al.*, 2009). Surprisingly, forced expression of key fate determining transcription factors is not only sufficient to perturb these networks, but also to successfully convert one cell type into the other, a process referred to as transdifferentiation or direct reprogramming (Iwafuchi-Doi and Zaret, 2014; Morris, 2016). Direct reprogramming was first described when fibroblasts were transdifferentiated into contracting muscle cells (Davis, Weintraub and Lassar, 1987; Weintraub *et al.*, 1989) and was soon extended to other paradigms, including direct reprogramming of ectoderm-derived glia into neurons (Heins *et al.*, 2002). Since then, reprogramming has been established for many cell types and even across germ layers, such as conversion of fibroblasts into neurons, hepatocytes and, most strikingly, induced pluripotent stem cells (Takahashi and Yamanaka, 2006; Vierbuchen *et al.*, 2010; Huang *et al.*, 2011a).

The current wealth of different reprogramming paradigms suggests that virtually any cell might be converted into any other cell type. However, despite the variety of available reprogramming paradigms and efforts to investigate the underlying molecular mechanisms (Buganim *et al.*, 2012; Chronis *et al.*, 2017; Fu *et al.*, 2018; Dall'Agnese *et al.*, 2019; Velychko *et al.*, 2019; Cates *et al.*, 2021; Kempf *et al.*, 2021; Yagi *et al.*, 2021) several simple, yet fundamental, questions remain unanswered, which likely limits our abilities to implement its full potential. For example, it is currently unclear whether different fate conversions share similar principles i.e., whether a general mechanism for starter cell identity loss exists. Also, it is unclear how much the expression levels of reprogramming factors influence both loss and gain of cell identity during reprogramming. Finally, another widely overlooked aspect of cell fate conversion is that it always entails a cell identity conflict. We know, however, very little about to what extent cells can resolve these and in which way they might fail. A major obstacle to address these questions from publicly available data is the use of optimized medium conditions during different reprogramming protocols, imposing a bias on gene expression (Chen *et al.*, 2014; Kim *et al.*, 2015; Kleijkers *et al.*, 2015; Ledur *et al.*, 2017). Hence, no direct comparisons of the impact various fate-instructing reprogramming factors have on starting cell identity have been performed to date. As a result, the question whether cells from the same germ layer are more readily converted into another as compared to conversion across germ layers also remains unanswered. Similarly, addressing whether expression levels of reprogramming factors are significant for converting the cellular transcriptome has been hindered by the predominant use of viral reprogramming vectors, which fail to be captured and quantified by single cell analysis.

Consequently, it remains unclear whether transcriptomic changes scale with reprogramming factor expression levels and, as a result, fate conversion occurs in a gradual manner; or whether instead a critical threshold exists. This might be particularly relevant for members of larger transcription factor families, such as basic helix loop helix (bHLH) proteins, which share low affinity sites and consequently are subject to binding site competition (Long, Prescott and Wysocka, 2016). Finally, although combinations of complementary factors have been used to obtain cell types of interest (Takahashi and Yamanaka, 2006; Vierbuchen *et al.*, 2010; Huang *et al.*, 2011b; Sekiya and Suzuki, 2011) it is unclear how cells would respond to conflicting fates being imposed on the cell at the same time.

To explore cell identity hierarchies and mechanisms of fate collision, we developed Collide-seq. This implements novel experimental approaches and analysis tools (fate titration) to quantify reprogramming factor levels and investigate, on the single cell level, how fibroblasts convert into lineages of different germ layers and how cell identity conflicts are resolved.

Results

A multiplexed strategy for the comparison and collision of cell fate conversions

To compare cell fate conversions and analyze the effect of their collisions, we established Collide-seq: a single cell protocol based on detectable and combinatorial reprogramming factor expression (**Figure 1B-C**). First, we selected a diverse panel of reprogramming factors: Ascl1, MyoD1, FoxA2, Sox2 and Pou5f1 (hereafter referred to as Oct4), known for driving differentiation towards different cellular identities belonging to distinct germ layers (e.g. neuronal, myogenic, hepatogenic, multipotent and pluripotent, **Table 1**) (Davis, Weintraub and Lassar, 1987; Takahashi and Yamanaka, 2006; Grinnell *et al.*, 2007; Vierbuchen *et al.*, 2010; Chen *et al.*, 2011; Huang *et al.*, 2011a; Li *et al.*, 2011; Tsai *et al.*, 2011; Wu *et al.*, 2011; Ring *et al.*, 2012; Chanda *et al.*, 2014; Lim *et al.*, 2016; Nakamori *et al.*, 2017). Importantly, the inclusion of Sox 2 and Oct4 allowed us to further investigate whether pluripotency factors act as general enablers of cell fate changes (Deleidi *et al.*, 2011; Kim *et al.*, 2011; Peskova *et al.*, 2019; Sharma *et al.*, 2019).

Next, we cloned these factors into a PiggyBac vector, placing their expression under the control of a doxycycline-dependent promoter (TRE) and added constitutively expressed fluorescent reporters (acGFP, EBFP2, DsRed) (**Figure 1B**). A key advantage of using PiggyBacs, over more commonly used viral approaches, is the generation of polyadenylated transcripts, facilitating transgene detection by scRNA-seq platforms (Mortazavi *et al.*, 2008). Typically, between 1 and 10 PiggyBac vectors integrate, leading to a broad range of transcript expression between different cells (Yusa *et al.*, 2009). Furthermore, the use of a doxycycline inducible promoter ensured synchronous expression, and a defined experimental starting point. Importantly, no protein expression was detected in the absence of doxycycline, but a strong nuclear signal was seen 48h after induction, indicating tight chemical control of reprogramming factor expression **(Figure 1D, Supplementary Figure 1A**). Constitutive fluorescent reporter expression cassettes enabled the enrichment and pooling of cells prior to transgene induction, i.e., initiation of cell fate conversion, and scRNA-seq.

To determine the consequences of inducing both the individual factors as well as up to three competing transcription factors simultaneously, we nucleofected mouse embryonic fibroblasts (MEFs) with the different factors as well as their combinations for a total of 17 different conditions, including cells expressing only the fluorescent reporters as a negative control (**Figure 1C**). After collecting an equal number of cells per condition and subsequently pooling them, cells were plated in a single well and transgene expression was induced for 72h prior to single-cell RNA sequencing (scRNA-seq) analysis using the 10x Genomics Chromium platform. In total, this yielded ~17.000 cells

(Replicate 1: 8709 cells, Replicate 2: 8219 cells) for downstream analysis after applying strict quality control criteria (see **Methods**). As expected, nearly all cells expressed fluorescent reporters as well as at least one reprogramming factor **(Figure 1E-F)**. Moreover, the expression of each transcription factors was localized to distinct parts of the *Uniform Manifold Approximation and Projection* (UMAP)(McInnes, Healy and Melville, 2018) embedding, indicating that reprogramming factor expression and transcriptomic effects were well correlated (**Figure 1E-F**). These attained cell states were overall also highly similar between both technical replicates, demonstrating that Collide-seq yields robust results (**Figure 1G**). Furthermore, reprogramming factor transcripts were confirmed to be mostly derived from PiggyBacs, as a custom single nucleotide polymorphism (SNP) based alignment strategy attributed the majority of reprogramming factor UMIs to the transgenic allele (**Supplementary Figure 1B**, see **Methods**). In addition, the observed cellular heterogeneity was not a technical artifact originating from ambient effects, as shown by applying ambient RNA correction to this data (Young and Behjati, 2018) (**Supplementary Figure 1C-F,** see **Methods**). Finally, cell cycle analysis indicated that most cells expressing a reprogramming factor had exited cell cycle after 72h, irrespective of whether the expressed reprogramming factor ultimately steers towards a post-mitotic cell fate (Ascl1 & MyoD1), or not (FoxA2, Oct4 & Sox2) (**Figure 1H**). Altogether, Collide-seq allowed reading-out the multiplexed expression of reprogramming TFs and correlated well with induced transcriptomic effects of the factors, thus making it a suitable system to explore the general principles of direct reprogramming.

De-multiplexing combinatorial reprogramming factor expression reveals discrete induction of distinct lineages

Our next aim was to demultiplex the dataset into its individual conditions. Previous approaches for single-cell knock-out state inference (Dixit *et al.*, 2016) are liable to yield false discoveries that derive from the knock-out state being modelled as a mixture model of gene expression. Hence, we devised a novel computational approach to assign cells to their maximum *a posteriori* fit condition leveraging only transgene and fluorophore expression (**Figure 2A**, **Supplementary Figure 2A,** see **Methods**). This model was regularized by the set of conditions defined in the experimental set-up (**Figure 1C**) and the inferred conditions mapped to distinct parts of the cell state space (**Figure 2A**). Accordingly, Louvain clustering stratified these states into distinct clusters corresponding to neurogenic (blue shades) myogenic (yellow/orange shades), hepatogenic (pink/purple shades) and multipotent (vermillion/red shades) transcriptional changes (**Figure 2B**). The multipotent cluster, however, was mainly dependent on Sox2, as Oct4 positive cells were intermingled with the original fibroblast population (**Figure 2A-B**, fibroblasts, grey shades). Furthermore, the fibroblast population showed

four cluster, three of which were close to each other, potentially representing different cell cycle states (**Figure 1H**), while one was more separated and possibly contained damaged cells given the relatively low gene and read count numbers (**Figure 2B,** dark gray shade, **Supplementary Figure 2B**). Most importantly, several clusters contained cells expressing more than one factor, indicating that collision between factors yield distinct cell states at 72h post induction **(Figure 2A-B**). These collision states did not show increased fractions of mitochondrial reads, a hallmark of cellular stress (Ilicic *et al.*, 2016; Luecken and Theis, 2019), implying that these states are viable (**Supplementary Figure 2C**). Taken together, these results revealed distinct transcriptomic consequences induced by each factor and their combinations which are well detected and tolerated for at least 72h after transgene induction, thereby validating these single factors as sufficient to induce distinct lineages

Comparing principles of fate conversion between distinct lineages

To identify both common and unique properties of different fate conversion processes, we first focused on transcriptomic changes in cells expressing only a single reprogramming factor. To this end, we fit a generalized linear model to the cellular expression vectors as a function of transgene expression and computed the overall differential expression magnitude for each factor across all genes (see **Methods**). This revealed a clear hierarchy in terms of these magnitudes (**Figure 3A,** see **Methods**), which can be interpreted as the induced deviation from the original fibroblast identity by each factor, which was largest for MyoD1 while Ascl1, FoxA2 and Sox2 had smaller but comparably strong effects. In accordance with its mild impact on the transcriptome (**Figure 2A**) the vector magnitude of Oct4 was smallest (**Figure 3A**). Therefore, we concluded that Oct4 failed to significantly perturb cell identity on its own and thus focused on the four remaining factors. A main advantage of Collide-seq is that, due to the multiplexing of conditions in a single biological replicate, most confounding factors are excluded. Hence, we could separate this general differential expression magnitude into two parts: fate acquisition and identity erasure which could be systematically compared these between the different factors.

For fate acquisition, demultiplexing as well as Louvain clustering revealed that all four factors show a wide range of reprogramming factor expression with strong cell-to-cell variation (**Supplementary Figure 3A**). Nevertheless, when individually expressed, each of them induced a distinct transcriptomic state (**Figure 2A-B**) that mapped to a defined lineage resolved through a pseudotemporal coordinate derived from an individual RNA velocity model (Bergen *et al.*, 2020) (**Figure 3B**). Moreover, using published ChIP-seq data (Oki *et al.*, 2018), a substantial overlap between putative binding sites and genes induced in our dataset, particularly for Ascl1, MyoD1 and

FoxA2 was found (**Figure 3C**). Furthermore, each factor also induced a distinct transcriptomic signature, characterized by unique target gene expression (**Figure 3D**) and gene ontology (GO) enrichment analysis on upregulated genes further confirmed the induction of lineages specific to each factor (**Supplementary Figure 3B-E**). Overall, these results confirmed the competence of Ascl1, MyoD1, FoxA2 and Sox2 to trigger transcriptomic changes of cell fate conversion towards their respective lineages.

To measure loss of identity, we devised a fibroblast score based on a literature derived list of fibroblast marker genes (**Table 2,** see **Methods**). Applying this score to the single factor data showed that all factors cause fibroblast identity loss after 72hrs of reprogramming factor induction, albeit to different extents (**Figure 3E**). A more global analysis on significantly downregulated genes validated this observation, as genes encoding for extracellular matrix and cell periphery proteins, core parts of the fibroblast proteome, were enriched for all four factors (**Supplementary Figure 4A-D, Methods**). Interestingly, however, the downregulated genes were rarely shared among all factors, suggesting that there is no common gene set or mechanism through which fibroblast identity is erased (**Figure 3F**). Notably, we did find substantial pairwise overlap in downregulated gene sets (e.g., 23% Sox2-FoxA2, 21% FoxA2-MyoD1, 35% Ascl1-MyoD1), suggesting similar strategies for some factor pairs, while others possibly follow different repressive schemes (**Figure 3F**). In line with having the biggest impact on the fibroblast transcriptome, MyoD1 was most potent in repressing the fibroblast signature, reflected by the fact that most of its downregulated genes were not shared with the other factors (**Figure 3F**). Overall, this analysis revealed that few similarities exist between different reprogramming factors when it comes to achieving fibroblast identity repression.

Collision of factors reveals mostly antagonistic effects

Next, we asked whether multiple factors synergize or antagonize in fate erasure. Using the same fibroblast score as above, factor collisions were shown to have intricate effects on switching off fibroblast identity (**Figure 4A**). For instance, the combination of Ascl1 and Sox2 repressed fibroblast identity more efficiently than Ascl1 alone (**Figure 4A**). In contrast, cells expressing Ascl1 and MyoD1 maintained a much higher fibroblast score than those expressing MyoD1 alone, while neither Sox2 nor FoxA2 had such a negative impact on MyoD1 (**Figure 4A**). This reinforces the concept that different molecular mechanisms exist for cell identity removal and demonstrates that these mechanisms can act synergistically in some instances, yet antagonistic or neutral outcomes are most common (**Figure 4A**, **Supplementary Figure 5A**). Furthermore, the fact that significant synergistic effects were only seen for Sox2 and Ascl1 indicates that the transcriptomic interactions follow

physiological principles, as both are expressed in neuronal progenitor cells (Pevny and Nicolis, 2010; Kageyama, Shimojo and Ohtsuka, 2019) and their co-expression efficiently reprograms human pericytes into neurons (Karow *et al.*, 2012, 2018). The addition of Sox2 or Oct4, on the other hand, did not improve identity loss for any of the other factors, suggesting that these pluripotency genes do not act as general enablers by facilitating starter cell identity removal, as suggested before (Deleidi *et al.*, 2011; Kim *et al.*, 2011; Peskova *et al.*, 2019; Sharma *et al.*, 2019).

To get a more global overview of the similarities across the different collisions, we analyzed how transcriptomic changes caused by one reprogramming factor are perturbed by the other factors. For this, we collapsed the cells within each condition into pseudobulk samples and performed hierarchical clustering (**Supplementary Figure 5B,** see **Methods**). Next, we performed principal component analysis of these pseudobulk samples and overlaid it with the hierarchical cluster labels (**Figure 4B, Supplementary Figure 5B,** see **Methods**). This revealed that no factor was clearly dominant over all other factors, although in some of the combinations one reprograming factor had significant dominance over the other (**Figure 4B**). Strikingly, MyoD1 only expressing cells were significantly different from all other conditions (**Figure 4B, Supplementary Figure 5B-C**). This unique transcriptional identity of the MyoD1 pseudobulk sample supports our findings on its differential expression magnitude (**Figure 3A**) and its striking potency in erasing fibroblast identity (**Figure 3E**), indicating MyoD1 is the putative most potent reprogramming factor in our setting. Nevertheless, the dissimilarity between double and triple factor combinations containing MyoD1 and the MyoD1 single factor condition indicated that the other factors can perturb the myogenic trajectory significantly (**Figure 4B**).

To quantify the degree of synergism and antagonism in collisions, we analyzed pair-wise interactions between factors in a linear model of cellular expression vectors. The interaction coefficients in this model can be interpreted as positive (synergistic) or negative (antagonistic) effects that are evoked by the collision on the expression vector of the single factor conditions. We correlated the coefficient vectors of each term in the linear model to understand their relative effect on gene expression (**Figure 4C**). For the individual reprogramming factors, mostly positive correlation coefficients were found, indicating a sign change in a similar direction (**Figure 4C**, upper left), likely caused by similarities in fibroblast identity removal (**Figure 3E-F**). In contrast, the effects observed upon the co-expression of two or more reprogramming factors produced a generally negative or neutral correlation, thus indicating an antagonistic or non-substantial change (**Figure 4C**, blue shades). This demonstrated that fate collisions are often detrimental to the individual fate, as

conversions are partially undone by the collision. Notably, pluripotency factors Oct4 and Sox2 were not found to act as general enablers of conversion in any of the combinations. In terms of strength, the negative correlation for the collision between MyoD1 and Ascl1 was strongest. As single factors, their effects showed a high correlation coefficient (**Figure 4C**, intersection of MyoD1 and Ascl1 and vice versa) confirming commonalities in their induced programs which have been reported previously (Lee *et al.*, 2020). However, their co-expression did not result in synergism, as might be expected, but rather displayed a very strong antagonistic effect (**Figure 4C**, intersection of MyoD1 and Ascl1 & MyoD1). Although perturbations could be seen for many other factors as well, they were clearly most prominent for the MyoD1 program. This pronounced sensitivity of the MyoD1 induced program was also seen when marker genes were examined, as expression of known MyoD1 targets was strongly perturbed by the other factors (**Figure 4D**). In contrast, Ascl1, Sox2 and FoxA2 retained the potency to express marker genes of their respective linage during fate collision better **(Figure 4D**).

Fate titration analysis reveals determinism of cell states driven by reprogramming factor levels

We speculated that the relative expression levels of the transcription factors may influence the degree of perturbation. Thus, we devised "Fate Titration Analysis" (FTA) to attribute transcriptomic states to reprogramming factor levels, modelling each cell as its own perturbation experiment with unique reprogramming factor expression levels and cell-state readout (see **Methods**). Transcriptomic states were represented by a clustering label and this grouping was used to assign cells to either of the two opposing fates or a collision thereof, which we referred to as a collision state (**Figure 4E**). Applying FTA to collisions involving MyoD1 consistently showed that cells expressing very low levels of the perturbing factor converted to the myogenic transcriptional program (**Figure 4F**). However, even at relatively low levels of expression, the perturbing factor redirected the transcriptome to a collision state that was markedly different from the two individual fates and in some cases cells were even directed to the state of the perturbing factor (Ascl1: Neurogenic, FoxA2: Hepatogenic, Sox2: Multipotent) (**Figure 4F**). This response seemed to be relatively independent of MyoD1 expression levels, as it occurred over the entire spectrum of MyoD1 expression (**Figure 4F**). Overall, these results indicated that the potential of reprogramming factors to induce global transcriptomic changes might be uncoupled from their sensitivity to perturbation.

The above FTA results also suggested that there may be determinism in gene expression states driven by reprogramming factor expression levels. One can think of determinism in the context of Waddington's landscape, describing not only the range of possible cell states, but also their relative stability as a function of reprogramming factor expression levels. We characterized the degree to which cell states can be predicted in our data with classification models of cell state, such as logistic regression of cluster assignments (see **Methods**). Indeed, cluster assignment could be fit with 0.70 accuracy (0.96 ROC AUC), showing that we could predict the general transcriptomic state of a cell based on its reprogramming factor expression levels, thus demonstrating determinism in the system at the level of cluster identities (**Figure 4G**). A stronger demonstration of determinism is the ability of a mathematical model to extrapolate transcriptomic states to unseen conditions. We performed this extrapolation on the cell-wise RNA vectors and were able to explain more variance with a nonlinear model than with a linear model (R2 of 0.61 versus 0.65, **Figure 4H**). Hence, outcome cell states are non-linear functions of reprogramming factor expression levels. This notion fits well to the above-described fate collisions which yield cell states in combinatorial reprogramming that are not simply the sum of the induced fates. We further hypothesized that the transcription factor combinations did not only yield a predictable cell state for a given transgene expression level, but also resulted in predictable cell state stability. Here, stability can be measured as gene expression variance where a low variance implies that a given transgene configuration instructs a very stable outcome cell state. Indeed, we were able to correlate the modeled gene-wise variance with prediction errors showing that variance estimation is possible here (R2=0.90) (**Supplementary Figure 5D**). When employing this, we found that variance does not increase with transgene levels, revealing that indeed each factor induces a stable transcriptomic program rather than a stochastic gene expression pattern (**Supplementary Figure 5E-G**). Lastly, we evaluated the ability of the identified rules of determinism to predict transcriptomic states in unseen genetic conditions (out-of-domain generalization). Expecting to generalizability of the fate decision-making rules learned by this model to increase with the complexity of the training data, we first predicted cell states in unseen triple-transgene conditions at an R2 of 0.63 (**Figure 4I**). Indeed, the predictive performance on all models was increased on triple-positive condition holdouts if the training data consisted of single- and double-positive conditions rather than single-positive conditions only (**Figure 4J**). Overall, this analysis suggested that the cell-wise gene expression distribution is globally deterministic in Collide-seq. The genetic circuits underlying fate decision making can be abstracted with supervised machine learning models as non-linear effects that reflect synergism and antagonism of factors. Taken together, the factor collisions shown here demonstrated that all tested somatic reprogramming factors possessed the potency to perturb the others, with no truly dominant factor. Furthermore, this disturbance was

highly deterministic as generalizable gene expression patterns could be learned with supervised machine learning models, indicating that model-aided design of reprogramming conditions for desired outcome fates is possible if enough gene expression data is provided.

Fate competition between the bHLH factors MyoD1 and Ascl1

Although each collision state is potentially different, we deemed following up the strong antagonistic interaction between MyoD1 and Ascl1 to be particularly interesting for several reasons. First, Ascl1 and MyoD1 are both bHLH factors that share several binding sites and interactors (Lee *et al.*, 2020; Martin, Sodaei and Santpere, 2021). Second, our data indicated that Ascl1 does not only perturb the myogenic program induced by MyoD1, but also impairs fibroblast identity loss induced by the myogenic factor (**Figure 4A**). We thus set out to study their competition in more detail by setting up a second Collide-seq experiment (**Supplementary Figure 6A**). For this experiment, we included a 24h and 48h timepoint, in addition to our original 72h timepoint, to get more insight into the temporal dynamics of fate collision. Furthermore, we chose to use Oct4 and Hnf1a, a fate determinant in liver development (Lau *et al.*, 2018), over FoxA2 and Sox2 as they were expected to be less potent transcription factors, allowing us to investigate with little confounding how a third fate factor would influence Ascl1 and MyoD1 competition. Applying the same computational approach as above demultiplexed the data into the different experimental conditions (**Supplementary Figure 6B**) and Louvain clustering stratified the different transcriptomic states (**Figure 5A**). As before, no unique clusters were found for Oct4 expressing cells at any of the included timepoints while Hnf1a only expressing cells formed a single distinct cluster (**Figure 5A, Supplementary Figure 6B**). In contrast, multiple distinct transcriptomic states for Ascl1 and MyoD1 only expressing cells could be detected and further confirmed by condition-wise RNA velocity analysis (**Figure 5A, Supplementary Figure 6C**). Moreover, a strong positive correlation between the velocity-based pseudotime and the time these three reprogramming factors were expressed demonstrated that duration of expression, rather than expression levels, moves fate conversions forward (**Supplementary Figure 6D**). Concentrating our analysis on Ascl1 and MyoD1 only expressing cells, these cells were found in two separate groups of adjacent clusters, referred to as neurogenic and myogenic clusters respectively (**Figure 5A**, Neurogenic: Blue shades, Myogenic: Yellow/Orange shades). Cells expressing Ascl1, together with either Hnf1a or Oct4, were mostly found in one of these neurogenic clusters (78% Ascl1 & Hnf1a, 84% Ascl1 & Oct4), while those that were positive for MyoD1 and either Hnf1a or Oct4 mostly resided in the myogenic clusters (75 % MyoD1 & Hnf1a, 84 % MyoD1 & Oct4) (**Figure 5A, Supplementary Figure 6B**). Cells expressing both Hnf1a and Oct4 mimicked the position of the Hnf1a single positives (**Figure 5A, Supplementary Figure 6B**). This indicated a clear hierarchy

between these four factors with Ascl1 and MyoD1 being most dominant, followed by Hnf1a, and then, lastly, Oct4.

Concentrating on the collision of the two most dominant factors, Ascl1 and MyoD1, we next computed lineage endpoints using the previously calculated RNA velocity vectors in combination with CellRank (Lange *et al.*, 2020), see **Methods**), grouping cells to shared fates in an unbiased manner. For Ascl1 and MyoD1 only expressing cells, a single unique lineage endpoint was found (**Figure 5B,** leftmost UMAPs). However, when combining both factors, two new collision states emerged which could be discretized into two Louvain clusters (**Figure 5B,** rightmost UMAP). Importantly, these transcriptomic states did not occur in the single transgene conditions and the predicted lineage endpoints were distinct from the single positive Ascl1 and MyoD1 endpoints (**Figure 5B**). Furthermore, *in silico* doublet simulation showed such cells would occupy very different transcriptomic states, suggesting that this collision state is not merely the average of the individual fates (**Supplementary Figure 6E-F).** We also confirmed that the occurrence of collision states between Ascl1 and MyoD1 is robust, as similar endpoints in a biological replicate dataset at 48h were found (**Supplementary Figure 7A-C**). Importantly, these collision endpoints were dominant over the endpoints for both individual factors, abolishing the purely neuronal attractor and only leaving few cells at the purely myogenic attractor (**Figure 5B**). Furthermore, based on their overall similarity to the respective single factor lineages, one collision state was found to be more neurogenic and the other more myogenic (**Supplementary Figure 6G**). Finally, overlapping our data with published ChIP-seq data (Lee *et al.*, 2020) confirmed that the co-expression of Ascl1 and MyoD1 results almost exclusively in antagonistic or neutral effects concerning target gene expression, in accordance with our above findings on factor collision and despite using a different computational approach (**Figure 4C, Figure 5C**).

The better temporal resolution and improved sampling of the entire transgene expression spectrum in this second Collide-seq experiment, allowed determining if and how levels of Ascl1 and MyoD1 expression influence the outcome of fate collision. To this end, we first looked at general target fate regulation, determining the general correlation between transgene and target gene expression levels by binning cells according to their reprogramming factor expression levels and examining the target gene expression for each bin (**Figure 5D**). Strikingly, a clear gene expression threshold at which both Ascl1 as well as MyoD1 either induced or repressed the expression of their target genes was found (**Figure 5D**). Hence, both the activation and repression of target genes seems to behave in a binary fashion, rather than a linear or exponential manner. Consequently, above this threshold

transcriptomic states were similar between cells within a single condition and changed little with additional increases in bHLH reprogramming factor level (**Figure 5D**).

Building on this, we then probed how expression levels of Ascl1 and MyoD1 influence the decision between collision states, again using FTA (see **Methods**). This showed that only cells with very high MyoD1 expression moved towards the myogenic collision or purely myogenic state (**Figure 5E**). At comparable levels of the respective transcription factors, cells mostly shifted into the more neurogenic collision state (**Figure 5E**). We quantified the degree of separation of both states with a linear classifier and were indeed able to distinguish cells from both states with high accuracy (0.72, **Methods**). This analysis illustrated the close call of fate competition by these very potent fate inducers: MyoD1 is dominant over Ascl1 at very high concentrations, but even low Ascl1 levels redirected a large proportion of cells to a more neurogenic state. This balanced cell fate competition is particularly intriguing, given that neurons are developmentally more distant from fibroblasts while muscle cells belong to the same germ layer.

Phenotypic analysis reveals DNA binding independence of the colliding factor

In the above-described results we restricted our analysis to transcriptomic data on fate collisions. However, an important read-out in reprogramming are changes in cellular morphology and function. For example, a major event in myogenesis is the fusion of newly formed myoblasts, giving rise to multi-nucleated myotubes (Sampath, Sampath and Millay, 2018). To investigate whether the collision between Ascl1 and MyoD1 affected such functional aspects of reprogramming, we nucleofected MEFs with either Ascl1, MyoD1, or their combination and performed immunocytochemistry at 3 days post transgene induction (**Figure 5F**). Indeed, induction of MyoD1 resulted in the generation of a substantial amount of multi-nucleated cells with myotube-like morphology and marker expression (**Figure 5F-G, Supplementary Figure 6H**). In contrast, the addition of Ascl1 resulted in a significant reduction of multinucleated muscle-like cell numbers (**Figure 5F-G**), as well as a marked decrease in the protein expression of muscle marker Desmin (**Supplementary Figure 6H**). To test whether this Ascl1-mediated perturbation of the myogenic program is a consequence of the collision of the two transcriptional programs or rather reflects competition between the two proteins themselves, we employed a mutant version of Ascl1 (mutAscl1) carrying two mutations in its basic domain: E131R132 to A131Q132 (see **Methods**). These residues are highly conserved in bHLH proteins and critical to DNA binding (Turner and Weintraub, 1994; Farah *et al.*, 2000). Consequently, mutAscl1 displayed a strongly reduced ability to activate the expression of direct Ascl1 targets (**Figure 5H**). Using mutAscl1 for fate collision with

MyoD1 showed that it was equally capable of reducing both the number of multinucleated cells as well as Desmin expression levels as wild type Ascl1 (**Figure 5F-G, Supplementary Figure 6H**). Overall, these results indicated that phenotypic perturbations observed upon the collision of Ascl1 and MyoD1 seems to be independent of Ascl1 DNA binding.

Competitive inhibition between Ascl1 and MyoD1 impairs pioneer factor activity

To further examine how Ascl1 and mutAscl1 perturb MyoD1 function on a molecular level, we set up a third Collide-seq experiment including these factors (**Figure 6A, Supplementary Figure 8A**). To get more insight into the underlying mechanisms, we performed joined single-cell profiling of gene expression and chromatin accessibility (see **Methods**). As before, we first confirmed the detection of transgene expression and demultiplexed the dataset into its individual conditions (**Figure 6B, 6F, Supplementary Figure 8B-C**). In agreement with the cell reprogramming assays (**Figure 5F**), collision with either Ascl1 or mutAscl1 caused a marked decrease in myogenic gene expression (**Figure 6C-D**). Supporting this notion, only very few colliding cells were found in the myogenic cluster (**Figure 6E**). These transcriptomic states were paralleled by condition-specific chromatin states, which also showed a disruption in double positives (**Figure 6F**). To further investigate whether this effect of cell fate collision is directly due to a disturbance of MyoD1 binding to its targets, we profiled DNA binding for MyoD1 using *cleavage under targets and release using nuclease* (CUT&RUN) (Skene and Henikoff, 2017) (**Figure 6A**). CUT&RUN for MyoD1 revealed an average of 7290 binding sites (**Figure 6G, Supplementary Figure 8D**). Fitting to the role of a developmental transcription factor, most of these binding sites were localized to gene regulatory elements, such as putative enhancer elements and gene promoters (**Figure 6H**). Strikingly, although the nature of bindings sites remained unchanged in collision conditions (**Figure 6H**), a vast reduction in DNA binding when MyoD1 collides with either wild-type or mutAscl1 was observed (**Figure 6G-H, Supplementary Figure 8D**). Indeed, coverage in the proximity of critical target genes revealed markedly reduced binding of MyoD1 upon collision with both Ascl1 and mutAscl1 (**Figure 6I, Supplementary Figure 8E**). As both Ascl1 and MyoD1 have been reported to act as pioneer transcription factors (Wapinski *et al.*, 2013, 2017; Casey *et al.*, 2018; Dall'Agnese *et al.*, 2019), we hypothesized that Ascl1 might hamper the pioneer activity of MyoD1. Overlapping our CUT&RUN data with our scATAC-seq and scRNA-seq data for these factors showed a near perfect coverage of CUT&RUN peaks by cumulative scATAC-seq peaks, confirming that the binding sites map to chromatin that is at least transiently open during the process (**Figure 6J, Supplementary Figure 8F**). Since the presence of both Ascl1 and mutAscl1 significantly decreased the general capability of MyoD1 to bind to and open chromatin (**Figure 6G, Supplementary Figure 8G**), which included several key myogenic lineage genes, we concluded that

they impact MyoD1's ability to act as a pioneer factor (**Figure 6K**). Although reduced in number, known motif analysis (see **Methods**) revealed that the fraction of binding sites with E-box motifs show a slight increase upon collision (**Figure 6L**). However, binding sites with the E-box motif of E2A (also known as Tcf3 or E47 and hereafter referred to as Tcf3), showed a substantial depletion (**Figure 6L).** Tcf3 is an E-protein and one of the main cofactors of MyoD1 and Ascl1 (Massari and Murre, 2000). Taken together, these results demonstrate that Ascl1 drives MyoD1 away from its binding sites and make it unable to induce the expression of myogenic target genes, thus resulting in a fate collision state. The reduced binding to sites containing a Tcf3 motif suggests that this may be caused by diminished availability of cofactors to MyoD1.

Discussion

Here, we have simultaneously compared and collided different cell fate conversions using Collide-seq (**Figure 7A**). Comparison of different reprogramming factors has, so far, mostly been performed with the aim of finding optimal conditions to drive cells towards a single given fate (Protze *et al.*, 2012; Yang *et al.*, 2019). However, to what extent different factors achieve cell conversion through similar or different mechanisms has received little attention. Collide-seq allowed exploring this question and revealed that there is a substantial difference in how different reprogramming factors erase the original identity (**Figure 3E**). This is interesting, because it indicates that, although quite stable under physiological conditions, several different entry points exist to erase the existing cell identity and characterizing and manipulating them might provide much needed strategies to improve existing reprogramming paradigms. Furthermore, fate erasure correlated well with factor potency overall, here quantified as perturbation magnitude in gene expression space with a linear model (**Figure 3A**). Interestingly, we found that the mesodermal factor MyoD1 was most potent in imposing its fate as well as repressing the original mesodermal fibroblast identity. This came as a surprise, because non-mesodermal factors were expected to silence more fibroblast genes rather than less and thus this result indicates that special relationships exist between starter cells and reprogramming factors, possibly related to germ layer identity (**Figure 3F**). Further evidence supporting this notion comes from MyoD1's inefficiency to transdifferentiate cells into myocytes when a starter cell of non-mesodermal origin is used (Davis, Weintraub and Lassar, 1987; Weintraub *et al.*, 1989). This might seem surprising, since not only Ascl1, FoxA2, Sox2 and Oct4 but also MyoD1 has been attributed pioneer factor activity (Iwafuchi-Doi and Zaret, 2014; Zaret and Mango, 2016; Zaret, 2020; Sunkel and Stanton, 2021). Therefore, none of these factors were predicted to be disturbed by the epigenomic setup or the identity of the starter cell. However, our results are in line

with the hypothesis that transdifferentiation into more closely related cell types might be more thorough and/or efficient (Hochedlinger and Plath, 2009; Morris and Daley, 2013). This suggests that therapeutic cell conversions may best be limited to related cell types if possible (Morris and Daley, 2013).

Interestingly, the strong reprogramming ability of MyoD1 on its own did not hold in competitive conditions. We discovered this surprising sensitivity of the pioneer factor through simultaneous induction of multiple divergent fates (Collide-seq). MyoD1's program was effectively perturbed by all other factors, except Oct4, and this effect was largely independent of the expression levels of MyoD1 and of the colliding factors (**Figure 4F**). Although we generally found antagonistic effects of colliding factors more frequently than synergism, one of the factors tested, Ascl1, was particularly disruptive and the only one capable of hindering MyoD1's potential to erase the fibroblast identity (**Figure 4A**). To our knowledge, these experiments are the first to systematically position different reprogramming factors against each other. Conceptually closest, perhaps, are cell fusion experiments (Cowan *et al.*, 2005; Brown and Fisher, 2021). Such experiments showed that pluripotency is often dominant over somatic cell identities. However, our fate collision showed no dominance of pluripotent programs, nor did we detect any hierarchy that would reflect a developmental order. This indicates that competition is probably resolved on the factor rather than the lineage level. Similarly, several studies have suggested that induction of pluripotency factors might be an effective way to remodel the epigenome and make differentiated cells more amenable to fate changes (Deleidi *et al.*, 2011; Kim *et al.*, 2011; Peskova *et al.*, 2019; Sharma *et al.*, 2019). However, our data show that Oct4 and Sox2 do not act as general cell fate enablers enhancing the transcriptional effects of the somatic reprogramming factors (**Figure 4C, Supplementary Figure 5A-B**). Instead, Oct4 affected the individual somatic reprogramming trajectories very little during fate collision. Notably, Oct4 has been suggested to remove cell identity by epigenetic mechanisms rather than inducing lineage specific gene expression, creating a more open and instable chromatin state that facilitates the actions of other factors (Kim *et al.*, 2011). It could be that this process is rather slow (Taberlay *et al.*, 2011), and hence not detectable during the first 72h examined here. Such a delayed effect of the pluripotency factors on the fibroblast chromatin might also explain why Sox2 and Oct4 double positive cells show only few regulated targets associated with pluripotency, although two-factor iPS reprogramming of fibroblasts has been conducted before (Huangfu *et al.*, 2008; Kim *et al.*, 2008; Nemajerova *et al.*, 2012).

Collide-seq enabled us to quantify the influence of reprogramming factor expression levels on cell conversion. Leveraging this with a machine learning approach we were able to reveal a high degree of determinism in this system by demonstrating that the induced perturbations yield predictable cell states. We also characterized transcription factor abundance-dependent fate choices with fate titration analysis, showing that Collide-seq can be used to titrate transcription factors against each other to screen fate outcomes. Interestingly, the resulting transcriptomic states were found to be a non-linear function of the reprogramming factor expression levels, which is both highlighted by the emergence of collision states in fate titration analyses as well as the step-like saturating dependency of gene activation of transcription factor induction ("binary switch"). In other words, cell fate conversion does not scale gradually with increasing reprogramming factor expression levels but instead depends on whether a critical expression threshold is reached. This finding indicates that and optimal level of reprogramming factors expression exists, which is sufficient for reprogramming but also minimizes the risk of failure due to cellular stress as a result of transcription factor overexpression. Thus, finding the optimal reprogramming factor expression levels could significantly improve existing reprogramming protocols. Furthermore, the cell-wise perturbation effects described here may also be leveraged for gene regulatory network modelling, as they tend to be stronger than available data based on single-cell knock out data. Hence, they may facilitate the development of bottom-up models for fate collision on the chromatin and transcriptome level in the future (Kamimoto, Hoffmann and Morris, 2020).

Overall, our study unraveled key principles of cellular reprogramming by systematically comparing reprogramming factors and applying collisions between them, thereby revealing a deterministic system that mapped the dominance or comparable strength of the different factors (**Figure 7A**). For example, the ability of Ascl1 to induce the neuronal lineage as well as remove fibroblast identity was similar to FoxA2 and Sox2 (**Figure 3A,E**). MyoD1, on the other hand, was found to be considerably more potent (**Figure 3A,E**)**.** However, this potency of MyoD1 was not reflected under competitive conditions, as their relative strengths reversed, making Ascl1 the strongest driver that directed the majority of cells towards a neurogenic intermediate **(Figure 5E)**. Correspondingly, we also observed a clear loss of MyoD1 pioneer activity upon collision with Ascl1 (**Figure 6K**). Together, these results revealed a profound effect of competition on MyoD1 function and prompted us to further investigate how Ascl1 achieves this. To this end, it is important to note that Ascl1 and MyoD1 are both basic helix loop-helix (bHLH) factors, which can function as either homo- or heterodimers (Massari and Murre, 2000; Wang and Baker, 2015; Murre, 2019; Martin, Sodaei and Santpere, 2021). Consequently, we considered four molecular mechanisms to explain these effects: i) collision between transcriptional programs, ii) competition between (shared) binding sites, iii) formation of

an inactive heterodimer and iv) cofactor competition (**Figure 7B**). Interestingly, we found that collision is hardly diminished by employing mutAscl1 (**Figure 5F, Supplementary Figure 6H**), indicating that fate collision is likely not a result of transcriptional competition or DNA binding site competition. An alternative explanation might therefore be that Ascl1 and MyoD1 form non-functional heterodimers that bind DNA but do not activate transcription. Our CUT&RUN data, however, shows, that many MyoD1 binding sites are lost upon collision with Ascl1, making this option also improbable. Thus, competition for possible dimerization partners appears to be the most plausible explanation for the observed effects. Indeed, our data revealed a reduced fraction of E2A/Tcf3 motifs in MyoD1 binding sites upon collision to support this (**Figure 6L**). In addition, a comparable interaction has already been suggested for MyoD1 and the bHLH factor Twist (Spicer *et al.*, 1996). The finding that the fate collision between MyoD1 and Ascl1 is triggered by competition between the factors themselves rather than their interaction with DNA or their programs is highly relevant for the design of novel reprogramming approaches, since it indicates that the master transcription factors expressed in the cell of origin rather than the molecular abilities of the expressed reprogramming factors might define, whether certain factors possess reprogramming potential in certain settings. It is important to note however that not all detected fate collisions might be caused by the same mechanism and that we unfortunately know very little about the factors that protect and define fibroblast identity. Taken together, we investigated fundamental concepts of cell identity conversions using a combination of novel machine learning and molecular biology approaches. These results exposed the underlying principles of cell identity acquisition that could be used to improve current reprogramming strategies through the informed selection of starter cells, cell fate factors and reprogramming factor expression levels.

Acknowledgements

We thank Tatiana Simon-Ebert, Judith Fischer-Sternjak and Christina Koupourtidou for technical help. We acknowledge the Core Facility Flow Cytometry at the Biomedical Center, Ludwig-Maximilians-Universität München and the Sequencing Core facility of the Helmholtz Institute Munich for providing equipment, services, and expertise. We thank Thomas Waltzhöni and Xavier Pastor Hostench for bioinformatics support provided at the Bioinformatics Core Facility, Institute of Computational Biology, Helmholtz Zentrum München. M.G. acknowledges funding by the German research foundation (SPP2306, SFB 870, SPP1757, Excellence Cluster SyNergy (EXC 2145 / Projekt-ID 390857198), the Roger de Spoelberch Foundation, and the EU (ERC advanced grants ChroNeuroRepair 340793, NSC Reconstruct 874758 and NeuroCentro). F.J.T. acknowledges support

by the BMBF (grant #L031L0214A, grant# 01IS18036A and grant# 01IS18053A), by the Helmholtz Association (Incubator grant sparse2big, grant # ZT-I-0007) and by the Chan Zuckerberg Initiative DAF (advised fund of Silicon Valley Community Foundation, 2018-182835 and 2019-207271). D.S.F. acknowledges support from the German Research Foundation (DFG) fellowship through the Graduate School of Quantitative Biosciences Munich (QBM) [GSC 1006 to D.S.F.] and by the Joachim Herz Stiftung.

Declaration of conflicts

F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH and Cellarity Inc., and ownership interest in Cellarity, Inc. and Dermagnostix. The authors have no competing interests or conflicts in relation to the manuscript.

Author contributions

S.H.S conceived this project. The experimental approach was designed by S.H.S and **B.A.H** together with M.G., F.J.T., G.M. and D.S.F. **B.A.H.** performed all cloning and biological experiments. T.W., D.R. and M.H. performed genome alignment for sequencing data. D.S.F. and K.M performed all bioinformatic analysis. S.H.S, D.S.F and B.A.H. wrote the manuscript and all authors contributed to corrections and comments.

References

Bergen, V. *et al.* (2020) “Generalizing RNA velocity to transient cell states through dynamical modeling,” *Nature Biotechnology*, pp. 1–7. doi:10.1038/s41587-020-0591-3.

Brown, K.E. and Fisher, A.G. (2021) “Reprogramming lineage identity through cell–cell fusion,” *Current Opinion in Genetics & Development*, 70, pp. 15–23. doi:10.1016/J.GDE.2021.04.004.

Buganim, Y. *et al.* (2012) “Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase.,” *Cell*, 150(6), pp. 1209–22. doi:10.1016/j.cell.2012.08.023.

Casey, B.H. *et al.* (2018) “Intrinsic DNA binding properties demonstrated for lineage-specifying basic helix-loop-helix transcription factors,” *Genome Research*, 28(4), pp. 484–496. doi:10.1101/GR.224360.117.

Cates, K. *et al.* (2021) “Deconstructing Stepwise Fate Conversion of Human Fibroblasts to Neurons by MicroRNAs,” *Cell Stem Cell*, 28(1), pp. 127-140.e9. doi:10.1016/J.STEM.2020.08.015.

Chanda, S. *et al.* (2014) “Generation of induced neuronal cells by the single reprogramming factor ASCL1,” *Stem Cell Reports*, 3(2), pp. 282–296. doi:10.1016/j.stemcr.2014.05.020.

Chen, Jiekai *et al.* (2011) “Rational optimization of reprogramming culture conditions for the generation of induced pluripotent stem cells with ultra-high efficiency and fast kinetics,” *Cell Research*, 21(6), pp. 884–894. doi:10.1038/cr.2011.51.

Chen, K.G. *et al.* (2014) “Human pluripotent stem cell culture: Considerations for maintenance, expansion, and therapeutics,” *Cell Stem Cell*. Cell Press, pp. 13–26. doi:10.1016/j.stem.2013.12.005.

Chronis, C. *et al.* (2017) “Cooperative Binding of Transcription Factors Orchestrates Reprogramming.,” *Cell*, 168(3), pp. 442-459.e20. doi:10.1016/j.cell.2016.12.016.

Cowan, C.A. *et al.* (2005) “Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells,” *Science*, 309(5739), pp. 1369–1373. doi:10.1126/SCIENCE.1116447.

Cunha, B.A., Domenico, P. and Cunha, C.B. (2000) “Pharmacodynamics of doxycycline,” *Clinical Microbiology and Infection*, 6(5), pp. 270–273. doi:10.1046/j.1469-0691.2000.00058-2.x.

Dall’Agnese, A. *et al.* (2019) “Transcription Factor-Directed Re-wiring of Chromatin Architecture for Somatic Cell Nuclear Reprogramming toward trans-Differentiation,” *Molecular Cell*, 76(3), pp. 453-472.e8. doi:10.1016/j.molcel.2019.07.036.

Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) “Expression of a single transfected cDNA converts fibroblasts to myoblasts,” *Cell*, 51(6), pp. 987–1000. doi:10.1016/0092-8674(87)90585-X.

Deleidi, M. *et al.* (2011) “Oct4-Induced Reprogramming Is Required for Adult Brain Neural Stem Cell Differentiation into Midbrain Dopaminergic Neurons,” *PLoS ONE*. Edited by J.C. Zheng, 6(5), p. e19926. doi:10.1371/journal.pone.0019926.

Enver, T. *et al.* (2009) “Stem Cell States, Fates, and the Rules of Attraction,” *Cell Stem Cell*. Cell Press, pp. 387–397. doi:10.1016/j.stem.2009.04.011.

Farah, M.H. *et al.* (2000) “Generation of neurons by transient expression of neural bHLH proteins in mammalian cells,” *Development*, 127(4), pp. 693–702. doi:10.1242/DEV.127.4.693.

Fu, K. *et al.* (2018) “Comparison of reprogramming factor targets reveals both species-specific and conserved mechanisms in early iPSC reprogramming.,” *BMC genomics*, 19(1), p. 956. doi:10.1186/s12864-018-5326-1.

GitHub - theislab/cellrank: Mapping the fate of single cells using RNA Velocity (2020). Available at: https://github.com/theislab/cellrank (Accessed: September 8, 2020).

GitHub - theislab/diffxpy: Differential expression analysis for single-cell RNA-seq data. (no date).

Grinnell, K.L. *et al.* (2007) “De-differentiation of mouse interfollicular keratinocytes by the embryonic transcription factor Oct-4,” *Journal of Investigative Dermatology*, 127(2), pp. 372–380. doi:10.1038/sj.jid.5700531.

Heinrich, C. *et al.* (2010) “Directing astroglia from the cerebral cortex into subtype specific functional neurons,” *PLoS Biology*, 8(5). doi:10.1371/journal.pbio.1000373.

Heins, N. *et al.* (2002) “Glial cells generate neurons: The role of the transcription factor Pax6,” *Nature Neuroscience*, 5(4), pp. 308–315. doi:10.1038/nn828.

Heinz, S. *et al.* (2010) “Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities,” *Molecular cell*, 38(4), pp. 576–589. doi:10.1016/J.MOLCEL.2010.05.004.

Hochedlinger, K. and Plath, K. (2009) “Epigenetic reprogramming and induced pluripotency,” *Development (Cambridge, England)*, 136(4), pp. 509–523. doi:10.1242/DEV.020867.

Huang, P. *et al.* (2011a) “Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors,” *Nature*, 475. doi:10.1038/nature10116.

Huang, P. *et al.* (2011b) “Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors,” *Nature*, 475(7356), pp. 386–391. doi:10.1038/nature10116.

Huangfu, D. *et al.* (2008) “Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2,” *Nature Biotechnology 2008 26:11*, 26(11), pp. 1269–1275. doi:10.1038/nbt.1502.

Ilicic, T. *et al.* (2016) “Classification of low quality cells from single-cell RNA-seq data.,” *Genome biology*, 17(1), p. 29. doi:10.1186/s13059-016-0888-1.

Iwafuchi-Doi, M. and Zaret, K.S. (2014) “Pioneer transcription factors in cell reprogramming,” *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 2679–2692. doi:10.1101/gad.253443.114.

Kageyama, R., Shimojo, H. and Ohtsuka, T. (2019) “Dynamic control of neural stem cells by bHLH factors,” *Neuroscience Research*, 138, pp. 12–18. doi:10.1016/J.NEURES.2018.09.005.

Kamimoto, K., Hoffmann, C.M. and Morris, S.A. (2020) “CellOracle: Dissecting cell identity via network inference and in silico gene perturbation,” *bioRxiv*, p. 2020.02.17.947416. doi:10.1101/2020.02.17.947416.

Karow, M. *et al.* (2012) “Reprogramming of pericyte-derived cells of the adult human brain into induced neuronal cells.,” *Cell stem cell*, 11(4), pp. 471–6. doi:10.1016/j.stem.2012.07.007.

Karow, M. *et al.* (2018) “Direct pericyte-to-neuron reprogramming via unfolding of a neural stem cell-like program,” *Nature Neuroscience 2018 21:7*, 21(7), pp. 932–940. doi:10.1038/s41593-018-0168-3.

Kempf, J. *et al.* (2021) “Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2,” *Cell Reports*, 36(7), pp. 109409–109410. doi:10.1016/J.CELREP.2021.109571.

Kim, J. *et al.* (2011) “Direct reprogramming of mouse fibroblasts to neural progenitors,” *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), pp. 7838–7843. doi:10.1073/pnas.1103113108.

Kim, J.B. *et al.* (2008) “Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors,” *Nature 2008 454:7204*, 454(7204), pp. 646–650. doi:10.1038/nature07061.

Kim, S.W. *et al.* (2015) “Modulation of the cancer cell transcriptome by culture media formulations and cell density,” *International Journal of Oncology*, 46(5), pp. 2067–2075. doi:10.3892/ijo.2015.2930.

Kleijkers, S.H.M. *et al.* (2015) “Differences in gene expression profiles between human preimplantation embryos cultured in two different IVF culture media,” *Human Reproduction*, 30(10), pp. 2303–2311. doi:10.1093/humrep/dev179.

Lange, M. *et al.* (2020) “CellRank for directed single-cell fate mapping,” *bioRxiv*, p. 2020.10.19.345983. doi:10.1101/2020.10.19.345983.

Lau, H.H. *et al.* (2018) “The molecular functions of hepatocyte nuclear factors – In and beyond the liver.” doi:10.1016/j.jhep.2017.11.026.

Ledur, P.F. *et al.* (2017) “Culture conditions defining glioblastoma cells behavior: what is the impact for novel discoveries?,” *Oncotarget*, 8(40), pp. 69185–69197. doi:10.18632/oncotarget.20193.

Lee, Q.Y. *et al.* (2020) “Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes,” *Nature Cell Biology*, 22(4), pp. 401–411. doi:10.1038/s41556-020-0490-3.

Li, H. *et al.* (2009) “The Sequence Alignment/Map format and SAMtools,” *Bioinformatics*, 25(16), pp. 2078–2079. doi:10.1093/BIOINFORMATICS/BTP352.

Li, Y. *et al.* (2011) “Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules,” *Cell Research*, 21(1), pp. 196–204. doi:10.1038/cr.2010.142.

Lim, K.T. *et al.* (2016) “Small Molecules Facilitate Single Factor-Mediated Hepatic Reprogramming,” *Cell Reports*, 15(4), pp. 814–829. doi:10.1016/j.celrep.2016.03.071.

Livak, K.J. and Schmittgen, T.D. (2001) “Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2 C T Method,” *METHODS*, 25, pp. 402–408. doi:10.1006/meth.2001.1262.

Long, H.K., Prescott, S.L. and Wysocka, J. (2016) “Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution,” *Cell*. Cell Press, pp. 1170–1187. doi:10.1016/j.cell.2016.09.018.

Luecken, M.D. and Theis, F.J. (2019) “Current best practices in single-cell RNA-seq analysis: a tutorial,” *Molecular Systems Biology*, 15(6), p. e8746. doi:10.15252/MSB.20188746.

la Manno, G. *et al.* (2018) “RNA velocity of single cells,” *Nature*, 560(7719), pp. 494–498. doi:10.1038/s41586-018-0414-6.

Martin, X. de, Sodaei, R. and Santpere, G. (2021) “Mechanisms of Binding Specificity among bHLH Transcription Factors,” *International Journal of Molecular Sciences 2021, Vol. 22, Page 9150*, 22(17), p. 9150. doi:10.3390/IJMS22179150.

Massari, M.E. and Murre, C. (2000) “Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms,” *Molecular and Cellular Biology*, 20(2), pp. 429–440. doi:10.1128/MCB.20.2.429-440.2000.

McInnes, L., Healy, J. and Melville, J. (2018) “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.” Available at: https://arxiv.org/abs/1802.03426v3 (Accessed: December 11, 2021).

Morris, S.A. (2016) “Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks,” *Development*, 143(15), pp. 2696–2705. doi:10.1242/dev.138263.

Morris, S.A. and Daley, G.Q. (2013) “A blueprint for engineering cell fate: current technologies to reprogram cell identity,” *Cell Research 2013 23:1*, 23(1), pp. 33–48. doi:10.1038/cr.2013.1.

Mortazavi, A. *et al.* (2008) “Mapping and quantifying mammalian transcriptomes by RNA-Seq,” *Nature Methods*, 5(7), pp. 621–628. doi:10.1038/nmeth.1226.

Murre, C. (2019) “Helix–loop–helix proteins and the advent of cellular diversity: 30 years of discovery,” *Genes & Development*, 33(1–2), pp. 6–25. doi:10.1101/GAD.320663.118.

Nakamori, D. *et al.* (2017) “Direct conversion of human fibroblasts into hepatocyte-like cells by ATF5, PROX1, FOXA2, FOXA3, and HNF4A transduction,” *Scientific Reports*, 7(1), p. 16675. doi:10.1038/s41598-017-16856-7.

Nemajerova, A. *et al.* (2012) “Two-factor reprogramming of somatic cells to pluripotent stem cells reveals partial functional redundancy of Sox2 and Klf4,” *Cell Death and Differentiation*, 19(8), pp. 1268–1276. doi:10.1038/cdd.2012.45.

Oki, S. *et al.* (2018) “ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data,” *EMBO reports*, 19(12), p. e46255. doi:10.15252/EMBR.201846255.

Peskova, L. *et al.* (2019) “Oct4-mediated reprogramming induces embryonic-like microRNA expression signatures in human fibroblasts,” *Scientific Reports*, 9(1), pp. 1–13. doi:10.1038/s41598-019-52294-3.

Pevny, L.H. and Nicolis, S.K. (2010) “Sox2 roles in neural stem cells,” *The International Journal of Biochemistry & Cell Biology*, 42(3), pp. 421–424. doi:10.1016/J.BIOCEL.2009.08.018.

Protze, S. *et al.* (2012) “A new approach to transcription factor screening for reprogramming of fibroblasts to cardiomyocyte-like cells,” *Journal of Molecular and Cellular Cardiology*, 53(3), pp. 323–332. doi:10.1016/J.YJMCC.2012.04.010.

Raudvere, U. *et al.* (2019) “g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update),” *Nucleic Acids Research*, 47(W1), pp. W191–W198. doi:10.1093/NAR/GKZ369.

Ring, K.L. *et al.* (2012) “Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor,” *Cell Stem Cell*, 11(1), pp. 100–109. doi:10.1016/j.stem.2012.05.018.

Robinson, J.T. *et al.* (2011) “Integrative genomics viewer,” *Nature Biotechnology 2011 29:1*, 29(1), pp. 24–26. doi:10.1038/nbt.1754.

Sampath, Srihari C, Sampath, Srinath C and Millay, D.P. (2018) “Myoblast fusion confusion: the resolution begins.,” *Skeletal muscle*, 8(1), p. 3. doi:10.1186/s13395-017-0149-3.

Schindelin, J. *et al.* (2012) “Fiji: an open-source platform for biological-image analysis,” *Nature Methods 2012 9:7*, 9(7), pp. 676–682. doi:10.1038/nmeth.2019.

Sekiya, S. and Suzuki, A. (2011) “Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors,” *Nature*, 475. doi:10.1038/nature10263.

Sharma, P. *et al.* (2019) “Oct4 mediates Müller glia reprogramming and cell cycle exit during retina regeneration in zebrafish,” *Life Science Alliance*, 2(5). doi:10.26508/lsa.201900548.

Skene, P.J. and Henikoff, S. (2017) “An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites,” *eLife*, 6. doi:10.7554/ELIFE.21856.

Spicer, D.B. *et al.* (1996) “Inhibition of Myogenic bHLH and MEF2 Transcription Factors by the bHLH Protein Twist,” *Science*, 272(5267), pp. 1476–1480. doi:10.1126/SCIENCE.272.5267.1476.

Sunkel, B.D. and Stanton, B.Z. (2021) “Pioneer factors in development and cancer,” *iScience*, 24(10), p. 103132. doi:10.1016/J.ISCI.2021.103132.

Taberlay, P.C. *et al.* (2011) “Polycomb-Repressed Genes Have Permissive Enhancers that Initiate Reprogramming,” *Cell*, 147(6), pp. 1283–1294. doi:10.1016/J.CELL.2011.10.040.

Takahashi, K. and Yamanaka, S. (2006) “Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors,” *Cell*, 126(4), pp. 663–676. doi:10.1016/j.cell.2006.07.024.

Tsai, S.-Y. *et al.* (2011) “Single Transcription Factor Reprogramming of Hair Follicle Dermal Papilla Cells to Induced Pluripotent Stem Cells,” *STEM CELLS*, 29(6), pp. 964–971. doi:10.1002/stem.649.

Turner, D.L. and Weintraub, H. (1994) “Expression of achaete-scute homolog 3 in Xenopus embryos converts ectodermal cells to a neural fate,” *Genes & development*, 8(12), pp. 1434–1447. doi:10.1101/GAD.8.12.1434.

Velychko, S. *et al.* (2019) “Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs,” *Cell Stem Cell*, 25(6), pp. 737-753.e4. doi:10.1016/J.STEM.2019.10.002.

Vickaryous, M.K. and Hall, B.K. (2006) “Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest,” *Biological Reviews of the Cambridge Philosophical Society*. Cambridge University Press, pp. 425–455. doi:10.1017/S1464793106007068.

Vierbuchen, T. *et al.* (2010) “Direct conversion of fibroblasts to functional neurons by defined factors,” *Nature*, 463(7284), pp. 1035–1041. doi:10.1038/nature08797.

Wang, L.H. and Baker, N.E. (2015) “E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease,” *Developmental Cell*. Cell Press, pp. 269–280. doi:10.1016/j.devcel.2015.10.019.

Wapinski, O.L. *et al.* (2013) “Hierarchical mechanisms for transcription factor-mediated reprogramming of fibroblasts to neurons,” *Cell*, 155(3). doi:10.1016/j.cell.2013.09.028.

Wapinski, O.L. *et al.* (2017) “Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons,” *Cell Reports* [Preprint]. doi:10.1016/j.celrep.2017.09.011.

Weintraub, H. *et al.* (1989) “Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD (muscle regulatory gene/MyoD retrovirus),” *Developmental Biology*, 86, pp. 5434–5438. Available at: http://www.pnas.org/content/pnas/86/14/5434.full.pdf (Accessed: February 19, 2018).

Wolf, F.A., Angerer, P. and Theis, F.J. (2018) “SCANPY: Large-scale single-cell gene expression data analysis,” *Genome Biology* [Preprint]. doi:10.1186/s13059-017-1382-0.

Wu, T. *et al.* (2011) “Reprogramming of Trophoblast Stem Cells into Pluripotent Stem Cells by Oct4,” *STEM CELLS*, 29(5), pp. 755–763. doi:10.1002/stem.617.

Yagi, M. *et al.* (2021) “Dissecting dual roles of MyoD during lineage conversion to mature myocytes and myogenic stem cells,” *Genes and Development*, 35(17–18), pp. 1209–1228. doi:10.1101/GAD.348678.121/-/DC1.

Yang, J. *et al.* (2019) “Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming,” *Stem Cell Reports*, 12(4), pp. 757–771. doi:10.1016/J.STEMCR.2019.02.010.

Young, M.D. and Behjati, S. (2018) “SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data,” *bioRxiv*, 2020, p. 303727. doi:10.1101/303727.

Yusa, K. *et al.* (2009) “Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon,” *Nature Methods*, 6(5), pp. 363–369. doi:10.1038/nmeth.1323.

Zaret, K.S. (2020) “Pioneer Transcription Factors Initiating Gene Network Changes,” *Annual Review of Genetics*, 54, pp. 367–385. doi:10.1146/ANNUREV-GENET-030220-015007.

Zaret, K.S. and Mango, S.E. (2016) “Pioneer transcription factors, chromatin dynamics, and cell fate control,” *Current Opinion in Genetics & Development*, 37, pp. 76–81. doi:10.1016/J.GDE.2015.12.003.

Code availability

We supply all notebooks used for data analysis available as Supplementary data 1 and publish generalizable code related to analyzing single cell reprogramming data as a python package ("tftools", Supplementary data 2, to be published on Github).

Data availability

The UMI count data of the single-cell experiments is supplied in Supplementary Data 3 for review and will be published on GEO.

Methods

Animals

R26-M2rtTA knock-in mice were obtained from Jackson Laboratory ((#006965) and maintained in pathogen-free conditions with 12h light/dark cycles. Mice were kept as homozygous for the knock-in.

Mouse embryonic fibroblasts isolation and culture

Mouse embryonic fibroblasts were obtained from E14.5 embryos of R26-M2rtTA knock-in mice. Using a dissection microscope (Leica), heads, limbs, vertebral column as well as internal organs were removed to make certain no multipotent cells were present in cultures. After dissection, 2-3 embryos were pooled, and tissue was dissociated in 0.15 % Trypsin (Gibco) for 10-15 minutes to obtain single cell suspensions. Cells were plated in one T75 tissue culture flask per embryo in MEF medium at 37°C and 5% CO_2 (Dulbecco's Modified Eagle Medium (Gibco #61965) supplemented with 10 % FBS (Pan Biotech #30-3302), 1% Sodium Pyruvate (Gibco), 1% HEPES (Gibco) and 1% Penicillin/Streptomycin). Cells were split once at a 1:3 ratio when confluent before freezing. After thawing, cells were grown in T75 culture flasks until confluent before nucleofection.

Nucleofection

For nucleofections, 1 μg of DNA per transcription factor and 2 μg of transposase construct was used in all conditions. Nucleofections were performed according to manufacturer's instructions (Lonza, P3 Primary Cell 4D-Nucleofector™ X). Briefly, 5.0 x 10^5 cells were counted and spun down at 200 rcf for 10 min, supernatant was discarded, and cells were resuspended in 100 μl nucleofection solution master mix (82 μl P3 Primary Cell Nucleofector Solution + 18 μl Supplement 1). Upon resuspension, cells were immediately transferred to previously prepared 1.5 ml tubes containing DNA mixtures and

subsequently transferred to 100 µl Nucleocuvettes. Nucleofection was performed using the CZ-167 program 500 µl of pre-warmed RPMI 1640 was added immediately after nucleofection (Gibco; 1 % Penicillin/Streptomycin). Cells were transferred to a 37°C and 5% CO_2 incubator for 10 min for recovery. Finally, cells were plated in 0.1% Gelatin in PBS (ROTI®Cell, Carl Roth) coated 12-well plates with 1 ml of preconditioned MEF medium and cultured at 37°C and 5% CO_2. For expansion, cells were subcultured on 0.1% Gelatin culture vessels at 37°C and 5% CO_2 until the appropiate cell number for the experiment was reached. For immunocytochemistry experiments cells were plated directly onto Poly-D-Lysine coated glass coverslips.

Piggy Bac Vector generation

To generate TetOn inducible Piggy Bac expression vectors for transgene expression, first a dCas9 expression cassette was removed from PB-dCas9-T2A-GFP-PolyA-Blasticidin (A. Köferle, unpublished, **Supplementary Data 4**) through cutting with *SpeI* and *Bsu36I* (New England Biolabs). Next, a Tet Response Element (TRE) followed by a minimal CMV promoter was amplified from pLV-TetO-Oct4 (pLV-tetO-Oct4 was a gift from Konrad Hochedlinger, Addgene plasmid # 19766; http://n2t.net/addgene:19766; RRID:Addgene_19766) using the primers TetOn Fwd and TetOn Rev (**Table 5**). Combining insert and backbone at a ratio of 3:1 in combination with the Gibson Assembly Mastermix (NEB, E2611S) for 30 min at 50℃, PB TetO PolyA was generated. Subsequently, a CMV enhancer element and CMV promoter driving the expression of either acGFP1 or EBFP2 were amplified from AcGFP1-C1 (AcGFP1-C1, was a gift from Michael Davidson, Addgene plasmid #54607, http://n2t.net/addgene:54607; RRID: Addgene_54607) and EBFP2-C1 (EBFP2-C1 was a gift from Michael Davidson, Addgene plasmid #54665, http://n2t.net/addgene:54665 , RRID: Addgene_54665) using Colors Fwd in combination with acGFP Rev or EBFP2 Rev respectively (**Table 5**). Using the same Gibson Assembly procedure as described above, both inserts were combined with PB TetO PolyA to generate PB TetO acGFP PolyA and PB TetO EBFP2 PolyA. DsRed Express 2 was amplified from pCAG-Ascl1-IRES-DsRed (Heinrich *et al.*, 2010) using the primers DsRed Fwd and DsRed Rev, the CMV enhancer and promoter were amplified from EBFP2-C (described above) using the primers CMV Fwd and CMV Rev and combined with PB TetO PolyA to generate PB TetO DsRed PolyA (**Table 5**). Next, an SV40 polyadenylation cassette was amplified from the AcGFP1-C1 plasmid using the primers SV40 Fwd (DsRed) or SV40 Fwd (acGFP & EBFP2) with SV40 Rev and combined with all the respective fluorescent reporter backbones to generate PB TetO PolyA acGFP PolyA, PB TetO PolyA EBFP2 PolyA and PB TetO PolyA DsRed PolyA through Gibson assembly (**Table 5**). Finally, Ascl1 was amplified from pCAG-Ascl1-IRES-DsRed (Heinrich *et al.*, 2010), MyoD1 was amplified from pCAG-MyoD1-IRES-GFP (in house), Oct4 was amplified from pLV-TetO-Oct4 (see above), Sox2 was amplified from pCAG-

Sox2-IRES-GFP (in house), FoxA2 was amplified from pLV-PGK-FoxA2 (pLV.PGK.mFoxa2 was a gift from Malin Parmar, Addgene plasmid # 33014 ; http://n2t.net/addgene:33014 ; RRID:Addgene_33014) and Hnf1a was amplified from E14.5 mouse liver cDNA. E14.5 mouse liver cDNA was obtained by tissue digestion in TRIzol (Invitrogen) according to manufacturer's instructions and subsequently performing reverse transcription of 100 ng input total mRNA using the Maxima First Strand cDNA Synthesis Kit (Thermo Fisher Scientific)(**Table 5**). The inserts of the respective factors were combined with all three fluorescent reporter carrying PiggyBac backbones, pre-digested with *Mfe1* (New England Biolabs), to generate PB TetO TF PolyA acGFP/EBFP2/DsRed PolyA, for all transcription factors using Gibson Assembly. For FoxA2 and Sox2, two AU rich elements were cloned in between the cDNA and polyA site by digesting both backbones with *Mlu1* (New England Biolabs) and combining it with an PCR amplified insert of AU-repetitive sequences (**Supplementary Data 4**) generated by amplification with AU Fwd and AU Rev from a custom DNA oligo (**Supplementary Data 4, Table 5**). To generate a 3xFLAG tagged MyoD1 construct, MyoD1 was amplified from PB TetO MyoD1 acGFP after digestion with *Hpa1* and *Kpn1* (New England Biolabs) using the MyoD1 FLAG Fwd and MyoD1 FLAG Rev primers (**Table 5**). A 3x FLAG repeat was amplified from a custom ordered DNA oligo (**Supplementary Data 4**) using the FLAG Fwd and Flag Rev primers (**Table 5**). Combining both inserts with the digested fluorescent reporter PiggyBac backbones mentioned above yielded PB TetO 3xFLAG MyoD1 PolyA acGFP/EFBP2/DsRed PolyA. To generate mutAscl1, point mutations E131R132 to A131Q132 were introduced in PB TetO Ascl1 acGFP using the site directed mutagenesis kit (New England Biolabs). All cloning primer sequences are provided in **Table 5** and all plasmid sequences of final constructs used in this study are provided in **Supplementary Data 4.**

Transgene induction

Transgene expression was induced by administration of doxycycline (2 µg/ml) every 24h until the indicated time points, considering its ± 22h half-life (Cunha, Domenico and Cunha, 2000). Culture medium was replaced by fresh doxycycline containing medium for each treatment. For immunocytochemistry experiments, transgene expression was induced they day after nucleofection and a non-treated sample of the same condition was used as a control. For single cell experiments and quantitative polymerase chain reaction (qPCR) experiments aimed at determining reprogramming factor expression levels, cells were first expanded to at T75-T175 before sorting (see below) and induction of transgene expression. A non-treated sample of the same condition was used as a control. For qPCR experiments aimed at quantification of Ascl1 and mutAscl1 downstream

targets, transgene expression was induced the day after nucleofection and a treated sample of untransfected cells was used as a control.

Immunocytochemistry & Image acquisition

For immunocytochemistry, cells were fixed at the indicated time points with 4% paraformaldehyde in phosphate buffered saline (PBS) for 10- 15 minutes. Cells were permeabilized by incubation with 3% BSA and 0.5% Triton-X 100 in PBS for 30 minutes. Primary antibodies were incubated in 3% BSA 0.5% Triton-X 100 in PBS at 4°C overnight or for 2 hours at room temperature (**Table 3**). Cells were thoroughly washed with PBS to rinse off the remaining primary antibody solution. Secondary antibodies (including DAPI) were incubated in the dark at room temperature for 1 hour in 3% BSA and 0.5% Triton-X 100 in PBS followed by thorough washing with PBS (**Table 3**). For cells plated on coverslips, coverslips were mounted on glass slides using a water based non-fluorescent mounting medium (Aqua Poly/Mount (Polysciences, Warrington, PA). Stained cells were analyzed using an AxioM2 or Axio Observer epifluorescence microscope (Carl Zeiss) for coverslips and culture plates respectively. Images were obtained using the ZEN2 Software (Carl Zeiss).

Fluorescence intensity quantification

To compare Desmin protein expression levels between conditions, images were acquired as described above using identical exposure times within an experiment. Intensities were quantified by loading images with Fiji (Schindelin *et al.*, 2012) and generate a mask for each cells using *Image > Adjust > Threshold*. Next, mean intensity within the masked area was measured using *Analyze > Measure*. For each image, a total of five equally sized regions without any clear Desmin signal were used to determine background levels and their mean was subtracted from the average measured intensity. To compare different experiments, intensity levels were normalized with respect to the MyoD1 only condition.

Fluorescence activated cell sorting

Cells were trypsinized, collected in pre-warmed MEF medium and washed once in PBS. Supernatant was discarded and cells were resuspended in 1 ml PBS supplemented with 10 % FBS before being transferred to FACS tubes by passing through a 40 µm cell strainer. Cells were sorted using a BD FACSARIAIIIu cell sorter (BD Biosciences). Gates were set using untransfected MEFs as a reference and further adjusted using transfected cells. Cells were sorted at flow rates between 2 and 3 (arbitrary units, corresponding to ~ 17-25 ul/min) and collected in 1.5 ml tubes containing 300 µl MEF medium supplemented with 10 % additional FBS. For quantification of reprogramming factor levels, 37.500 - 100.000 cells were sorted per condition and plated in a single well of a 24-well plate

before inducing transgene induction with doxycycline the next day (see above). For single cell experiments 7.000/10.000 cells for each transfected condition and 5.000/7.500 untransfected MEFs were sorted. After sorting, all cells were pooled in a single 15 ml conical tube and centrifuged at 300 rcf for 5 min. 1 ml of supernatant was left and supplemented with fresh MEF medium. Cells were plated by taking 1 ml of this suspension containing between 50-60.000 cells in a single 24-well. For quantification of Ascl1 and mutAscl1 target expression levels, 100.000 cells per condition were sorted 48h after induction and immediately processed for RNA isolation (see below).

RNA extraction and quantitative polymerase chain reaction (qPCR)

For RNA extraction, cells from a single 24-well were collected 48 hours after transgene induction and RNA was isolated using the ARCTURUS® PicoPure® RNA Isolation Kit (Applied Biosystems) according to manufacturer's instructions. Genomic DNA was removed using the On-Column DNase I Digestion Set (Sigma-Aldrich). For retrotranscription, 100 ng of total mRNA was retrotranscribed using the Maxima First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). First-strand cDNA was diluted 1:5 in RNAse free water and 5 µl was used for each qPCR reaction. qPCRs were performed on a QuantStudio 6 (Applied Biosystems) using PowerUp™ SYBR™ Green Master Mix (Applied Biosystems). The expression of each gene was determined in triplicate and relative expression determined using the ΔΔCt method (Livak and Schmittgen, 2001). Primers are listed in **Table 3**.

***In vitro* reprogramming**

For reprogramming experiments, cells were nucleofected as described above and expanded on 0.1% gelatin coated cover slips. After expansions, single (Ascl1, mAscl1 & MyoD1) as well as double positive cells (Ascl1 & MyoD1, mAscl1 & MyoD1) were sorted and plated at a density of ~12.500 cells per well in a 96-well plate. The next day, medium was changed and transgene expression induced as described above. Cells were treated with doxycycline every 24 hours for the first 4 days, after that they were treated every other day. Cells were fixed at 3 days post induction and immunochemistry was performed as described above.

Droplet based scRNA-seq and scATAC-seq

For scRNA-seq only samples, cells were trypsinized and collected in pre-warmed MEF medium at the indicated time points. Cells were centrifuged at 300 rcf for 5 min to remove debris and resuspended in 50 - 100 µl PBS for cell counting. Cell suspensions were counted, and suspension volume adjusted to contain approximately 1000 cells per µl. For a targeted retrieval of 10.000 cells, ~17.500 cells were loaded, and libraries were prepared using the Chromium Single Cell 3' Reagent Kits v2 (24, 48,

72h data with Ascl1, MyoD1, Hnf1a and Oct4) or v3 (72h data with Ascl1, MyoD1, FoxA2, Sox2 and Oct4) according to manufacturer's instructions. Libraries prepared with v2 chemistry were sequenced on a HiSeq4000 whereas v3 libraries were sequenced on a NovaSeq 6000. All libraries were sequenced with a 100 bp paired end configuration.

For scMultiome (scRNA-seq + ATAC-seq) samples, cells were trypsinized and collected in pre-warmed MEF medium 72 hours after induction. Cells suspension was centrifuged at 300 x g for 5 min at 4°C and resuspended in 50 µl of PBS + 0.04% BSA (Miltenyi Biotec). Cells were pelleted once more by centrifugation at 300 x g for 5 min at 4°C. 45 µl of supernatant was removed and an equal amount of lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% Nonidet P40 Substitute, 0.01% Digitonin, 1% BSA, 1 mM DTT, 1U/ul H_2O) was added. Cells were lysed on ice and after 5 min 50µl of wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 1% BSA, 0.1% Tween-20, 1 mM DTT, 1 U/ul RNAse inhibitor) was added without mixing. Nuclei were pelleted by centrifugation at 500 x g for 5 min at 4°C. 95 µl of supernatant was removed, taking care not to disturb the pellet, and 45 µl of diluted nuclei buffer (1x Nuclei Buffer (10x Genomics), 1mM DTT, 1 U/ul RNAse inhibitor) was added without mixing. Nuclei were spun down once more at 500 x g for 5 min at 4°C and all supernatant was removed. Nuclei were resuspended in 7 µl of ice-cold nuclei buffer and 2 µl of nuclei suspension was mixed with 8 µl of diluted nuclei buffer and 10 µl Trypan Blue to determine nuclei concentration. For a targeted retrieval of 10.000 nuclei, the nuclei suspension was diluted to a concentration between 3280 and 8060 nuclei per µl. Libraries were prepared using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit according to manufacturer's instructions and sequenced on an Illumina NovaSeq 6000 sequencer.

Cleavage Under Targets and Release Using Nuclease (CUT & RUN)

For CUT&RUN assays, 3xFLAG-MyoD1 (see above) was used to allow MyoD1 pulldown and DNA binding assessment. The assay was performed using the CUT&RUN assay kit (Cell Signaling Technologies, 86652) according to manufacturer's instructions. Briefly, 100.000 cells per reaction collected and bound to Concanavalin A Magnetic beads. Cells were permeabilized and incubated with a primary antibody against FLAG for 2 hours at 4°C. Subsequently, cells were incubated with pAG-MNase for 1 hour at 4°C. pAG-MNase was activated by adding calcium chloride and a 30 min incubation at 4°C. Stop buffer (Cell Signaling Technologies) with 10 pg of Spike-In DNA (Cell Signaling Technologies) was added to each sample to stop the reaction and obtain normalization reads after sequencing. DNA was purified using a phenol/chloroform extraction and ethanol precipitation as described in the manufacturer's protocol. Input samples were generated by collecting 100.000 cells per condition and incubating with DNA extraction buffer (Cell Signaling Technologies) at 55°C for 1

hour shaking at ~750 rpm on a ThermoMixer. Afterwards, samples were cooled down and sonicated with a BioRuptor® Pico (Diagenode) using 10 sets of 30-sec pulses. Fragmented chromatin was isolated together with CUT & RUN samples as described above.

CUT & RUN Library preparation

DNA sequencing libraries were generated using the SimpleChIP® ChIP-seq DNA Library Prep Kit for Illumina (Cell Signalling Technologies, 56795) and SimpleChIP® ChIP-seq Multiplex Oligos for Illumina® (Dual Index Primers, Cell Signaling Technologies, 46538) following the manufacturer's instructions and adapting the protocol where needed as instructed by the manufacturer in the CUT & RUN Assay kit protocol. Briefly, an equal amount of DNA was used for all CUT&RUN and input samples and DNA ends were prepared for adaptor ligation. Note that here incubation temperature was lowered from 65°C to 50°C according to the manufacturer's instructions. Next, adapters were ligated to the DNA. Finally, DNA was amplified using PCR and Dual Index primers for Illumina® (Cell Signaling technologies, 47538). Importantly, anneal and extension time was reduced from 75 seconds to 15 seconds to avoid amplification of large library fragments per the manufacturer's instructions. Furthermore, all clean-up steps were performed with 1.1x volume of SPRIselect® beads to increase capture of smaller DNA fragments. Generated libraries were pooled and sequenced using a the MiSeq Reagent Kit v3 and a 2 x 75 bp paired-end sequencing strategy on an Illumina® MiSeq sequencer.

Preprocessing of sequencing data

For the alignment of reads we used the sequences and annotation files for the Mouse genome (GRCm38) from Ensembl (release 97). The synthetic transcription factors and reporter sequences and a custom generated annotation were appended to the genome sequence and annotation. The Cell Ranger software (version 3.1.0) run with the command 'cellranger mkref' created an index of the genome. The Cell Ranger pipeline run with the command 'cellranger count' aligned the reads, generated QC metrics, estimated the number of valid barcodes and created the count matrices. The command was executed with standard parameters, except that we adjusted the number of expected cells and the chemistry parameter accordingly. AnnData objects of the raw and filtered count matrices were created using the Python package Scanpy ((Wolf, Angerer and Theis, 2018), version 1.4.4). The alignment files (bam) were sorted by barcode using samtools ((Li *et al.*, 2009), version 1.10) and Velocyto ((la Manno *et al.*, 2018), version 0.17.17) was run using the 'velocyto run' command with the filtered barcodes list from the Cell Ranger run.

Processing of CUT&RUN data

The CUT&RUN data was processed using the Nextflow based ChIP-seq workflow from nf-core (version 1.2.2). For the different samples, a single pseudo antibody was defined and the samples of the same conditions sequenced in multiple experiments were defined as replicates. The computational pipeline was run in 2 independent executions to align against the yeast genome (R64-1-1) and against the mouse genome (GRCm38). The aligned reads from the yeast alignment were used for normalization as per the manufacturer's instructions found here: https://www.cellsignal.com/learn-and-support/protocols/cut-and-run-protocol.

CUT&RUN peak visualization

For visualization of CUT&RUN peaks, CUT&RUN data was processed as described above. Processed data was loaded into Integrative Genomics Viewer (IGV, version 2.11.4)(Robinson *et al.*, 2011). To compare several samples, each sample was loaded as an individual track. For exporting images, y-axes were set to the same scale before exporting.

CUT&RUN Motif analysis

Motif analysis was performed using Homer v4.11(Heinz *et al.*, 2010). Enriched motifs were acquired by comparing peaks to the mouse mm10 genome using findMotifsGenome.pl <BED file> mm10 <output directory> -size 200.

Processing of single cell multiome data

The single cell Multiome-Seq data was processed using the software cellranger-arc (version 2.0.0), run with the cellranger-arc count command with standard parameters and the corresponding genome. To build the reference genome index for mouse, GRCm38 the annotation from Ensembl (release 97) was used. The sequences and custom annotations of the synthetic transcription factors and fluorophores were appended to the genome fasta and annotation files. The samples processed with cellranger-arc count were aggregated using the cellranger-arc aggr command with the --normalize=none parameter.

Generation of a modified gene annotation

The sequences of the overexpressed reprogramming factors differ from the endogenous loci in only a short interval in their UTRs. To distinguish between endogenous and transgene expression, we created synthetic versions of the endogenous genes by replacing their UTRs with the UTR sequences

of the transgenes (from Gencode version vM20, https://www.gencodegenes.org), and appended them to the gene annotation. This modified annotation was then used for read alignment by the Cell Ranger pipeline as described above (Preprocessing of sequencing data). The Cell Ranger pipeline only considers reads that are compatible with a single gene annotation, and therefore only reads specific to the UTRs of the endogenous genes or of the transgenes are used for UMI counting.

Computational demultiplexing

The exact analysis is documented in **Supplementary Data 1, Notebook 1**. We first grouped cells into induced and non-induced for each transgene and fluorophore by defining an expression threshold for each factor (maximum likelihood assignment of cells to centroids). Secondly, we assigned cells to the condition of their posterior probability, using the initial assignments to define the prior distribution of transgene counts in induced cells (**Supplementary Figure 2A, B**, the posterior membership probability). This second step allowed us to incorporate the prior knowledge that only a subset of all possible transgene combinations was included in this setup.

Firstly, we defined condition centroids with two separate models, that yielded similar results in the end. Firstly, we used a gradient-free optimization algorithm to infer optimal decision boundaries θ between the active and inactive component for each transgene in a joint optimisation problem, using the deviation between the inferred distribution of cells over conditions f and the input distribution *x* as an objective *l*:

$$l = |f(\theta) - x|$$

Here, the maximum likelihood assignment of a cell to a condition $\hat{c}$ consists of its classification based on the decision boundaries:

$$\hat{c} = 1 \; if \; \theta > x \; else \; 0$$

Secondly, we defined centroids with a 2-component Gaussian mixture model for each transgene separately. Here, the maximum likelihood assignment of a cell to a condition consists of its assignment to a mixture *k*:

$$\hat{c} = arg\max_k P(x|\mu_k, \sigma_k) = arg\max_k -\log(\sigma_k) - \log(2\pi) - \frac{1}{2}\frac{(x-\mu_k)^2}{{\sigma_k}^2}$$

These mixture models showed collapse of the inactive component to zero in a couple of cases, in which we then augmented the inferred threshold to achieve similar results to the gradient-free optimization method (**Supplementary Data 1, Notebook 1**).

Posterior membership probability: The maximum likelihood assignments proposed above cannot guard against assignment of cells into conditions that do not exist, such as a condition with all transcription factors active and all colors inactive. Due to noise in the data, such conditions may

indeed be the maximum likelihood estimator of condition membership. However, we have the prior knowledge that only a subset of all transgene conditions exists in the data. We defined a partition function containing only the input conditions $\mathbb{K}$ and defined a posterior probability distribution over input conditions using this partition function and Gaussian likelihoods *P* to then assign cells to their maximum a posteriori estimate of membership:

$$\hat{c} = arg\max_{k} \frac{P(x|\mu_k, \sigma_k)}{\sum_{k \subset \mathbb{K}} P(x|\mu_k, \sigma_k)}$$

Unsupervised analysis of single-cell RNA-seq data

The exact analysis is documented in code in **Supplementary Data 1, Notebook 1**. We removed cells with high mitochondrial content (>20% of UMIs are from mitochondrial genes) or low total mRNA counts (<1000 expressed genes). We then filtered non-highly variable genes, normalized the mRNA counts within each cell, log(x+1) transformed the counts, computed PCA with 50 components, computed a k-nearest neighbor graph with k=X and computed UMAP embeddings and Louvain clustering on the graph. We used SoupX (Young and Behjati, 2018) for ambient RNA correction (**Supplementary Data 1, Notebook 1b**). We defined the cell cycle score as a z-score over cells on the sum of G2M-phase score and S-phase score computed with scanpy (Wolf, Angerer and Theis, 2018), **Supplementary Data 1, Notebook 1**). We computed the fibroblast score as a z-score defined over cells in the Louvain clusters annotated as non-cycling fibroblasts on the sum of log-normalized fibroblast marker genes per cell: Thy1, Col1a1, Postn, Vim, Prrx1, Timp3, Ccn2, Col5a1, Col1a2, Glipr2, Itgb5, Sh3kbp1, Tex264, Tnc, Cnn1, Fn1, S100a4, Twist2, Snai2, Cav1, Ecm1, Acta2, Col4a1, Col5a2, Mmp2, Mmp14, Mmp23, Col3a1, Cav2, Timp1, Timp2, Fgf7, Vcl, Itgb8 (also defined in **Supplementary Data 1, Notebook 1**).

Cellular doublet simulation

We considered doublet simulation between Ascl1 single-positive cells and MyoD1 single-positive cells to establish whether this specific doublet could explain the Ascl1+MyoD1+ double-positive cellular state. As both source distributions are defined by cells from the respective single-positive conditions, we sampled (n=200) random pairs of simulated doublets, added their transcriptomic states (unprocessed counts) and divided these counts by two to receive simulated doublet cells in the range of total counts of the single-positive cells. We then performed an unsupervised analysis workflow as described in (Unsupervised analysis of single-cell RNA-seq data) on the union of these simulated cells and the real cells from all conditions (**Supplementary Figure 6E-F**). To map the simulated cells into the annotated clusters defined on the real cells, we used a nearest neighbor classifier, only using the real cells with the minimal distance observed to each simulated cell.

Differential expression analysis

The exact analysis is documented in **Supplementary Data 1, Notebook 3**. We used Wald tests on a negative binomial generalized linear model (*GitHub - theislab/diffxpy: Differential expression analysis for single-cell RNA-seq data.*, no date) fit on cell cycle, condition and normalization factors testing the condition effect to determine differentially expressed genes along the individual lineages (**Supplementary Data 1, Notebook 3**). To restrict the setting to each lineage, we only considered cells from the single-positive condition of the lineage and the empty vector control for the test. P-values were corrected for multiple testing using Benjamini-Hochberg correction. We declared genes as differentially expressed if they had a corrected p-values of less than 0.01, a fold change of less than 2/3 or more than 3/2 and a mean expression level of at least 0.05 counts. We added the label up- or down-regulated in multiple places for those differentially expressed genes with positive (up) and negative (down) log fold change. We inferred synergism and antagonism in a differential test on cells from the control condition, two considered single-positive conditions and the double-positive condition of these two, modelling each transgene's effect individually and their interaction and tested and reported the interaction effect to yield synergism and antagonism labels: We defined synergistic effects as positive transgene interaction effects in a differential expression model if the log fold change in the double positive condition was positive (a gene is upregulated more strongly in the double positive in the double-positive condition than would be expected based on the sum of both transgenes in their individual single-positive conditions). Additionally, we defined negative transgene interaction effects in a differential expression model as synergistic if the log fold change in the double positive condition was negative (stronger down-regulation than expected). Conversely, we defined weaker up-regulation and weaker down-regulation as antagonistic. When defining differential expression vectors, we used the L2 norm over all gene-wise parameter estimators for an individual coefficient defined in the model, only using genes for which this coefficient was significant (FDR-corrected p-values < 0.01) and which were not overfit (defined as a L1 norm of significant effects within a gene > 100).

Gene ontology analysis

Gene ontology enrichment analysis was performed using the g:Profiler Python client (Raudvere *et al.*, 2019). Differentially expressed genes (Up: log-2-foldchange > 2/3, Down: log-2-foldchange < -2/3, log-normalized mean expression > 0.1 and adjusted p-value < 0.01) for each condition were provided as input and all list of all expressed genes in the dataset were used as background. For the downregulated genes, gene ontology analysis was performed using the cellular component

annotations whereas for upregulated genes the annotation biological process was used. The top 20 enriched GO terms were ranked based on p-value and visualizes using a custom graphing function (see Supplementary Material 1).

ChIP-seq data analysis and synergism and antagonism annotation

The exact analysis is documented in **Supplementary Data 1, Notebook 3**. We used published peak files for both Ascl1 and MyoD1 ChIP (Lee *et al.*, 2020). We overlapped these peaks against 5' ends of gene annotations (GRCm38), extending the annotated interval 5' end 10kb upstream and 2kb downstream.

RNA velocity and Cellrank analysis

The exact analysis scvelo (Bergen *et al.*, 2020) and cellrank ((*GitHub - theislab/cellrank: Mapping the fate of single cells using RNA Velocity*, 2020)) worflows are documented in code in **Supplementary Data 1, Notebook 2**. The RNA velocities mentioned in the main text are based on the dynamical RNA velocity model implemented in scvelo. Additionally, we fit the dynamical model with lenient gene filtering and a steady-state model with standard gene filtering. We ran CellRank (Lange *et al.*, 2020) on the dynamical RNA velocity model fits separately for each inferred condition using the CFLARE model on a kernel consisting of 50% transcriptomic connectivity and 50% RNA velocities. The attractors were assigned to fates by cellrank and are colored according to the annotated Louvain clusters established in the unsupervised analysis section.

Neurogenic and myogenic scores

The exact analysis is documented in **Supplementary Data 1, Notebook 4**. We defined the neurogenic score as *1-diffusion pseudotime* from the most mature neurogenic cell: We defined the most mature neurogenic cell as the cell with the highest latent time assigned by scvelo during dynamical RNA velocity inference in the Ascl1 single-positive condition. The scored cells presented are all cells in neurogenic or myogenic states excluding the cycling cluster. Similarly, we defined the myogenic score within the MyoD1 single-positive condition. The myogenic score presented in **Supplementary Figure 6G** is computed as above but computed on the set excluding cell in the mature myogenic cluster to compare intermediate myogenic and neurogenic states.

Fate titration analysis

The exact analysis is documented in **Supplementary Data 1, Notebook 2**. We scaled transgene counts into an active range between the minimum transgene count observed in the positive

condition and the 99th percentile of transgene counts in the single positive condition (maximum): $x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$. We visually compared two groups of cells (cell rank attractor groups or Louvain cluster groups) using 2D kernel density estimators of the joint distribution of each group over both transgenes and using the respective marginal distributions of each group. We fit a weighted logistic regression model with class-balancing weights between both groups on the scaled transgene counts. We report the accuracy of this classifier on a held-out test partition of 20% of the data.

Supervised modelling

All model and analysis code is also documented in Suppl. Data 1,2. We used linear and nonlinear models both for classification of categorical cluster assignments and regressive prediction of the log-normalized RNA observations of each cell. We accounted for confounding in all models: We corrected for the sample through a one-hot encoded predictor and for the cell size through a log-transformed size factor in the input. In all cases, the input to these models is either the one-hot encoded condition assignment (categorical model, **Figure 4G-H**), the binary presence of transgenic transcription factors (binary models, **Figure 4G-H**), or the log transgenic transcription factor expression per cell (expression models, **Figure 4I-J**). The nonlinear model was a neural network of fully connected layers with two hidden layers of 64 units each and tanh activation function. We used 1e-6 L1 and L2 penalties on linear layers. The hyperparameters of all models are also described in the supplied code. We trained all networks until convergence and evaluated performance on 10% of randomly selected test cells (**Figure 4G-H**) or on entirely held-out cell sets corresponding to inferred conditions (**Figure 4I-J**). We trained on all cells that were not in the test set, except of in **Figure 4I** where we tested models trained on only single- or single- and double-positive conditions on selected triple-positive conditions. We used area under the curve of the receiver-operator characteristic (AUC ROC), accuracy, top-3 accuracy and condition class-balanced accuracy to evaluate classification models and used R2 per cell between observed and predicted log-normalized expression for regression models.

Figure legends

Figure 1 A multiplexed strategy for the comparison and collision of cell fate conversions

A. Illustration of Waddington's landscape depicting key factors during development, reprogramming and transdifferentiation **B.** Schematic outline of Collide-seq **C.** Overview of experimental conditions **D.** Representative immunofluorescent images of 48h after transgene induction. Scale bar represents 50 μm. **E.** Log-normalized fluorophore expression superimposed on Uniform Manifold Approximation and Projection (UMAP). Shown are total count-normalized fluorophore expression levels on an ln scale. **F.** Log-normalized transgene expression superimposed on UMAP. Shown are total count-normalized transgene expression levels on an ln scale. **G.** UMAP embedding of scRNA-seq dataset colored by technical replicate **H.** Cell cycle score superimposed on UMAP (see **Methods**).

Figure 2: De-multiplexing combinatorial reprogramming factor expression reveals discrete induction of distinct lineages

A. Visualization of assignment outcome (see **Methods)** superimposed on Uniform Manifold Approximation and Projection (UMAP). Depicted are the total population of cells carrying one, two or three factors (left, blue) as well as the individual conditions (right, red) for the indicated conditions. For individual conditions, numbers indicate total number of cells per condition **B.** Louvain clustering superimposed on UMAP (left panel) and separated for individual lineages (right panels). Shown are Louvain clusters named according to the lineage of their predominant transcription factor. Cells without clear transcriptomic changes compared to the control conditions were labeled as fibroblasts.

Figure 3: Comparing principles of fate conversion between distinct lineages

A. Bar plot of transgene-wise coefficient vector magnitudes. **B.** Superimposed on Uniform Manifold Approximation and Projection (UMAP) is the latent pseudotime (lineage progression coordinate) computed with scvelo (Bergen *et al.*, 2020) in the context of the velocity computation **C.** Confusion matrix showing the intersection of upregulated genes in individual factor conditions versus putative binding sites derived from a publicly available ChIP dataset (Oki *et al.*, 2018) **D.** Matrixplot showing the relative expression of key target genes between different single factor conditions and empty vector carrying fibroblasts **E.** Fibroblast score (see **Methods)** for cells single positive for a given transcription factor compared to fibroblasts. **F**. Venn diagram of downregulated gene intersection between single factor conditions.

Figure 4: Factor collision reveals antagonistic effects of a deterministic system

A. Fibroblast score (see **Methods)** for cells single and double positive for given transcription factor combinations in comparison to fibroblasts. **B.** PCA scatter plot of first (vertical) and second (horizontal) loading showing the different conditions after collapsing them into pseudobulk samples. Colored rectangles correspond to clusters defined in **Supplementary Figure 5B**. EV: Empty vector expressing fibroblasts **C.** Correlation matrix of linear model effects for each condition. Color defines direction of correlation (negative = blue, positive = red) and color tone depicts size of correlation. **D.** Stacked violin plots of lineage marker genes for each factor alone and when additional factors are expressed. **E.** Conceptual summary of cell assignment for fate titration analysis (left most panel) as well as data derived examples for the indicated combinations of factors (right panels) **F.** Fate titration plots of Ascl1, FoxA2 and Sox2 versus MyoD1. Shown are the decision boundaries for indicated intermediate fates according to transcription factor levels. Transcription factor expression shown on the x and y axis are log-normalized expression values scaled into the dynamic range of the single-positive condition (see **Methods**). **G,H**. Predictive performance for a linear model of the categorical condition variable, a linear model of the log of the transgene expression, and a non-linear model of the log of the transgene expression in randomly held-out test cells. For the task of Louvain cluster assignment prediction, shown are area under the receiver-operator characteristic curve (AUC ROC), top-3 accuracy, accuracy, and class-balanced accuracy (G). For the task of prediction of log-normalized expression values of highly variable genes, shown is the cell-wise R2 (G). **I,J.** Predictive performance on held-out reprogramming conditions for models and metric as described in (H), with the exception of the base line model, which is a linear model of binary transgene presence. The hold-out task in (I) is a particular triple-positive condition as indicated in the legend. The hold-out task in (J) is the set of all triple positive conditions and models are either trained on single-positive conditions only or on single- and double-positive conditions.

Figure 5: Fate competition between the bHLH factors MyoD1 and Ascl1

A. Louvain clustering superimposed on Uniform Manifold Approximation and Projection (UMAP) for second Collide-seq experiment **B.** Condition based RNA velocity from designated conditions. Shown is a UMAP of all cells in data set (grey). Superimposed are cells from the indicated conditions colored in with the terminal fate assigned by CellRank (Lange *et al.*, 2020), see **Methods**). **C.** Intersections of genes regulated by Ascl1 and/or MyoD1 with ChIP-seq data (Lee *et al.*, 2020), see **Methods**). Shown are gene numbers and percentages grouped by synergistic and antagonistic effects in expression. **D.** Expression Z-score of genes up- and downregulated by Ascl1 (left) and MyoD1 (right) versus transcription factor expression levels binned into 10% intervals. **E.** Fate titration plot of Ascl1 and

MyoD1. Shown are decision boundaries for indicated intermediate fates according to transcription factor levels. MyoD1 and Ascl1 expression shown on the x and y axis are log-normalized expression values scaled into the dynamic range of the single-positive condition (see **Methods**). **F.** Representative immunofluorescent images of cells 3 days after transgene induction. **G.** Quantification of cells (%) with more than one nuclei among all transfected cells within a given condition. Pairwise comparisons performed with Welch's t-test, correction for multiple testing performed with Benjamini-Hochberg correction. *: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **H.** Log fold change of target gene expression as measured by qPCR after 48h of induction with mutAscl1 (blue bars) or wild-type Ascl1 (orange bars). Fold change is relative to untransfected dox induced cells. Data represents n =1 biological replicate. Error bars reflects standard deviation between technical replicates.

Figure 6: Competitive inhibition between Ascl1 and MyoD1 impairs pioneer factor activity

A. Schematic overview of third Collide-seq experiment. **B.** Visualization of assignment outcome (see **Methods)** for individual cells superimposed on UMAP of scRNA-seq part of experiment. Depicted are cells positive for the indicated conditions using all cells (top panels) as well as the top 500 highest transgene expressing cells (bottom panels). Bottom right panel depicts grouped Louvain clustering based on the predominant condition present **C.** Matrixplot showing the relative expression of key myogenic genes between different clusters as indicated in panel B. Myo = Myogenic, Col = Collision, Con = Control **D.** Log2 fold change of myogenic gene expression in double positive conditions as compared to MyoD1 only expressing cells. **E.** Fraction of cells of per condition amongst the different Louvain clusters **F.** Visualization of assignment outcome (see **Methods)** for individual cells superimposed on UMAP of scATAC-seq part of experiment. Depicted are cells positive for the indicated conditions using all cells (top panels) as well as the top 500 highest transgene expressing cells (bottom panels). Bottom right panel depicts grouped Louvain clustering based on the predominant condition present **G.** Confusion matrix of averaged CUT&RUN peaks for individual conditions for a total of 2 independent biological replicates. **H.** CUT&RUN peak classification averaged between 2 independent biological replicates depicted as fraction of total peaks. **I.** Representative Integrative Genome Browser (IGV, see **Methods)** (Robinson *et al.*, 2011) tracks for the indicated samples in CUT&RUN sample 1 out of 2. **J.** Bar plot showing the number of averaged CUT&RUN peaks over two independent biological replicates that are covered by scATAC-seq. **K.** CUT&RUN (left panel), scATAC (middle panel) and scRNA-seq (right panel) signal for key myogenic marker genes. CUT&RUN signal is averaged for two biological replicates **L.** Bar plot showing the percentage of MyoD1 binding sites that contain the canonical E-box motif (blue) as well as the

E2A/Tcf3 motif (orange). For E-box motifs, all variations were grouped and averaged within each condition.

Figure 7: Proposed model

A. Schematic summary of factor hierarchies deducted from our experiments. Strength of interaction between factors is reflected by line width. Nature of interaction (occurrence of collision state/dominance) is depicted by arrow heads (collision state) or bar ends (dominance).
B. Schematic overview of the possible mechanisms for Ascl1 and MyoD1 collision. i) collision between transcriptional programs, ii) competition between (shared) binding sites, iii) formation of an inactive heterodimer and iv) cofactor competition.

Supplementary Figure 1

A. Quantification of transgene induction using qPCR. Shown is the mean fold change of doxycycline induced samples over non-induced samples after 48 hours (n = 2/3). **B.** Fraction of transgenic UMIs successfully mapped to transgenic allele per population of cells expressing a single factor. **C.** Uniform Manifold Approximation and Projection (UMAP) embedding of scRNA-seq data colored by technical replicate after ambient RNA correction. **D.** Cell cycle score superimposed on UMAP for ambient corrected data. **E.** Log-normalized fluorophore expression superimposed on UMAP after ambient RNA correction. Shown are total count-normalized fluorophore expression levels on an ln scale. **F.** Log-normalized transgene expression superimposed on UMAP after ambient RNA correction. Shown are total count-normalized transgene expression levels on an ln scale.

Supplementary Figure 2

A. Assigned conditions in first, maximum likelihood, step of algorithm for a gradient-free optimization method by sample and by transgene (see **Methods**). **B.** Average number of genes (top panel) and UMI counts (bottom panel) for selected Louvain clusters. **C.** Fraction of mitochondrial UMIs per condition.

Supplementary Figure 3

A. Distribution of log-normalized transcription factor expression levels in single positive conditions (Ascl1: light blue, MyoD1: orange, FoxA2:pink, Sox2: vermillion and Oct4:blue) as well as empty vector controls (yellow). **B-E.** Top 20 terms related to biological process after gene enrichment (GO) analysis using G-profiler (see **Methods**) for genes upregulated by Ascl1 (B), MyoD1 (C), FoxA2 (D) and

Sox2 (E). Gene overlap with each term is depicted as the size of the circles and color reflects the adjusted p-value. On the y-axis the gene ratio is shown.

Supplementary Figure 4

A-D. Top 20 terms related to cellular component after gene enrichment (GO) analysis using G-profiler (see **Methods**) for genes downregulated by Ascl1 (A), MyoD1 (B), FoxA2 (C) and Sox2 (D). Gene overlap with each term is depicted as the size of the circles (bigger is more) and color reflects the adjusted p-value. On the y-axis the gene ratio is shown.

Supplementary Figure 5

A. Fibroblast score (see **Methods)** for indicated conditions. **B.** Similarity of pseudobulk principal components per condition clustered with hierarchical clustering. Colors correspond to those used in Figure 4B. **C.** Correlation matrix of linear model effects showing for correlation between individual factors and collision state. Color defines direction of correlation (negative = blue, positive = red) and color tone depicts size of correlation **D.** Correlation of predictive accuracy, measured as explained variance, on held-out test data with the predicted uncertainty of the model, measured as the mean of the predicted standard deviation by gene. **E.** Explained variance of model prediction by condition and replicate. **F,G.** Explained variance of cells binned by individual transgene expression (**F**) and individually with trend fit (**G**). In G, the superimposed color is the predicted uncertainty measured as the mean standard deviation over output genes.

Supplementary Figure 6

A. Schematic overview of experimental conditions **B.** Visualization of assignment outcome (see **Methods)** for individual cells superimposed on Uniform Manifold Approximation and Projection (UMAP). Depicted are cells positive for the indicated conditions as well as their total number. **C.** Condition based RNA velocity trajectories. Shown are RNA velocities computed within single positive cells, projected on a UMAP and computed on the transcriptomes of all cells (see **Methods).** Superimposed in color is the latent pseudotime (lineage progression coordinate) computed with scvelo (Bergen *et al.*, 2020) in the context of the velocity computation. **D.** Relative abundance of cells from each collection time point per cluster for Ascl1 (left) and MyoD1 (right). **E.** Fraction of cells per Louvain cluster colored according to double positive class: *in silico* simulated or actual double positives. **F.** Simulated double positives and actual double positives colored and superimposed on UMAP. **G.** Neuronal and myogenic score for Ascl1 & MyoD1 double positives cells residing in

collision states. **H.** Normalized Desmin fluorescence intensity for indicated conditions (left panel) as well as representative examples (right micrographs). Scale bars represents 100 µm.

Supplementary Figure 7

A. Log-normalized transgene expression superimposed on Uniform Manifold Approximation and Projection (UMAP). Shown are total count-normalized transgene expression levels on ln scale. **B.** Louvain clustering superimposed on UMAP. **C.** Condition based RNA velocity on a UMAP computed on cells from designated conditions. Shown is a UMAP of all cells in data set (grey). Superimposed are cells from the indicated condition colored in with the terminal fate assigned by CellRank (Lange *et al.*, 2020), see **Methods**).

Supplementary Figure 8

A. Schematic overview of experimental conditions. **B.** Log-normalized transgene expression superimposed on UMAP. Shown are total count-normalized transgene expression levels on an ln scale. **C.** Visualization of assignment outcome (see **Methods)** for individual cells superimposed on scRNA-seq Uniform Manifold Approximation and Projection (UMAP, top panels) and scATAC-seq UMAP (bottom panel). Depicted are cells positive for the indicated conditions. **D.** Confusion matrix of CUT&RUN peaks for individual conditions of CUT&RUN experiments 1 and 2 **E.** Representative Integrative Genome Browser (IGV, see **Methods)** tracks for the indicated samples in CUT&RUN sample 2 out of 2 **F.** Bar plot showing the number of CUT&RUN peaks covered by scATAC-seq for CUT&RUN experiments 1 and 2**. G.** Confusion matrix showing the number of ATAC peaks in the different experimental conditions.

Tables

Table 1: Overview of different reprogramming factor and the characteristics of their target cells

TF	Target cell	Germ layer	Potency	Cell cycle status
Ascl1	Neuron	Ectoderm	Terminally differentiated	Post mitotic
MyoD1	Skeletal muscle	Mesoderm	Terminally differentiated	Post mitotic
FoxA2	Hepatocyte	Endoderm	Terminally differentiated	Mitotic
Sox2	Neural stem cells	Ectoderm	Multipotent	Mitotic
Oct4	Pluripotent stem cell	Pre-germ layer	Pluripotent	Highly proliferative

Table 2: Fibroblast score genes

Thy1	Ccn2	Tex264	Snai1
Col1a1	Col5a1	Tnc	Cav1
Postn	Col1a2	Cnn1	Ecm1
Vim	Glipr2	Fn1	Acta2
Prrx1	Itgb5	S100a4	Col4a1
Timp3	Sh3kbp1	Twist2	Col5a2
Mmp2	Mmp14	Mmp23	Col3a1
Cav2	Timp1	Timp2	Fgf7
Vcl	Itgb8		

Table 3: qPCR primers

Target	Species	Forward primer (5’ – 3’)	Reverse primer (5’ – 3’)
Ascl1	Rat	TCG GCT ACA GCC TTC CAC	CAA AGC CCA GGT TAA CCA AC
MyoD1	Mouse	TGG CAT GAT GGA TTA CAG CGG	TCA CTG TAG TAG GCG GTG TC
Hnf1a	Mouse	GAG AAA CGC GTG GCT CTG AA	GTC CTC CTG AAG AAG TGA CTC CAC

Oct4	Mouse	AGC ACG AGT GGA AAG CAA CT	CTT TCA TGT CCT GGG ACT CCT C
FoxA2	Mouse	CGA GCA CCA TTA CGC C	GGG TAG TGC ATG ACC TG
Sox2	Mouse	CGA GCT GGT CAT GGA GTT GTA C	AAC GGC AGC TAC AGC ATG ATG C
Hes6	Mouse	ACG GAT CAA CGA GAG TCT TCA	TTC TCT AGC TTG GCC TGC AC
Dll1	Mouse	GGG ACA GAG GGG AGA AGA TG	CAC ACC CTG GCA GAC AGA T
Pvalb	Mouse	GGC AAG ATT GGG GTT GAA G	AGC AGT CAG CGC CAC TTA G
Cbfa2t3	Mouse	TGGCTACATGCCTGAAGAGAT	GCGCTTCACCTCATTCACA
GAPDH	Mouse	TTG CAG TGG CAA AGT GGA GA	CGT TGA ATT TGC CGT GAG TG

Table 4: Antibodies

Primary Antibody	**Host species**	**Isotype**	**Dilution**	**Source**
RFP	Rabbit	-	1:1000	Rockland (600-401-379)
GFP	Chicken	-	1:1000	Aves Labs (GFP-1020)
Ascl1	Mouse	IgG1	1:400	BD Pharmingen (556604)
MyoD1	Mouse	IgG1	1:500	BD Pharmingen (554130)
Hnf1a	Rabbit	-	1:400	Cell Signaling Technology (89670S)
Oct4	Rabbit	-	1:250	Abcam (181557)
Sox2	Mouse	IgG1	1:400	Abcam (79351)
FoxA2	Rabbit	-	1:400	Cell Signaling Technology (8186S)
FLAG	Mouse	IgG1	1:200	Sigma-Aldrich (F1804)
Secondary antibody		-		
DAPI	Goat	-	1:1000	
Anti-chicken 488	Goat	-	1:1000	ThermoFisher A11039
Anti-rabbit 546	Goat	-	1:1000	ThermoFisher A21207
Anti-mouse IgG1 594	Goat	-	1:1000	ThermoFisher A21125

Table 5: Cloning primers & Oligos

Primer name	**Primer sequence (5’ – 3’)**
TetOn Fwd	GGC GCA GTA GTC CAA ACA GGG ACA GCA GAG ATC

TetOn Rev	TAG GCA GCC TGC ACC TGA GGA GCT CGA GAG GTC AGG TCA AAA CAG CGT GGA TG
Colors Fwd	ACG CTG TTT TGA CCT CAC TAG TCG TTA CAT AAC TTA CGG TAA
acGFP Rev	CTG CAC CTG AGG AGC TTA CTT GTA CAG CTC ATC C
EBFP2 Rev	CTG CAC CTG AGG AGC TTA CTT GTA CAG CTC GTC C
CMV Fwd	CTG TTT TGA CCT GAC CTC ACT AGT CGT TAC ATA ACT TAC GGT AAA TG
CMV Rev	ATG GTG GCG AGT CCG GTA GCG CTA GC
DsRed Fwd	CCG GAC TCG CCA CCA TGG CCT CCT CC
DsRed Rev	CTG CAC CTG AGG AGC CTA CAG GAA CAG GTG GTG GCG
SV40 Fwd (Red)	CTG TTT TGA CCT GAC CTC ACT GAA ACA TAA AAT GAA TGC AAT TG
SV40 (Green & Blue)	CGC TGT TTT GAC CTC ACT GAA ACA TAA AAT GAA TGC AAT TG
SV40 Rev	CCG TAA GTT ATG TAA CGA TAA GAT ACA TTG ATG AGT TTG G
Ascl1 Fwd	TGA AAC ATA AAA TGA ATG CCT AAG CCA CCA TGG AGA GCT CTG GCA AGA TG
Ascl1 Rev	TAA ACA AGT TAA CAA CAA CTC AGA ACC AGT TGG TAA AG
MyoD1 Fwd	TGA AAC ATA AAA TGA ATG CGC CAC CAT GGA GCT TCT ATC GCC G
MyoD1 Rev	TAA ACA AGT TAA CAA CAA CAT CTC TCA AAG CAC CTG ATA AAT C
Hnf1a Fwd	TGA AAC ATA AAA TGA ATG CGC CAC CAT GGT TTC TAA GCT GAG CC
Hnf1a Rev	TAA ACA AGT TAA CAA CAA CGG CAC TTA CTG GGA AGA GGA GGC
Oct4 Fwd	TGA AAC ATA AAA TGA ATG CGC CAC CAT GGC TGG ACA CCT GGC TTC
Oct4 Rev	TAA ACA AGT TAA CAA CAA CGT GCC TCA GTT TGA ATG CAT GGG
AU Fwd	ATG TAA GTC GAG TTG TTG TT
AU Rev	TAA GCT GCA ATA AAC AAG TT

MyoD1 FLAG Fwd	CGA TGA CAA GAT GGA GCT TCT ATC GCC G
MyoD1 FLAG Rev	CGA TGA CAA GAT GGA GCT TCT ATC GCC G
FLAG Fwd	CTG AAA CAT AAA ATG AAT GCT GCT GGG CCA CCA TGG ACT
FLAG Rev	GAA GCT CCAT CTT GTC ATC GTC ATC CTT GTA ATC G

Supplementary Data

Extended Data 1: Analysis notebooks detailing single-cell RNA-seq analysis. Some analyses presented here depend on the newly developed python packages provided in Supp. Data 2.

Download link: https://hmgubox2.helmholtz-muenchen.de/index.php/s/dnmFRDqz4bsaogq
Password: P774Qm6wfw

Extended Data 2: tftools python package to run custom analysis centered on transcription factors used in this code.

Download link: https://hmgubox2.helmholtz-muenchen.de/index.php/s/dnmFRDqz4bsaogq
Password: P774Qm6wfw

Extended Data 3: Alignment output of single-cell experiments.

Download link: https://hmgubox2.helmholtz-muenchen.de/index.php/s/dnmFRDqz4bsaogq
Password: P774Qm6wfw

Extended Data 4: PiggyBac construct sequences

Download link: https://hmgubox2.helmholtz-muenchen.de/index.php/s/dnmFRDqz4bsaogq
Password: P774Qm6wfw

Figure 1

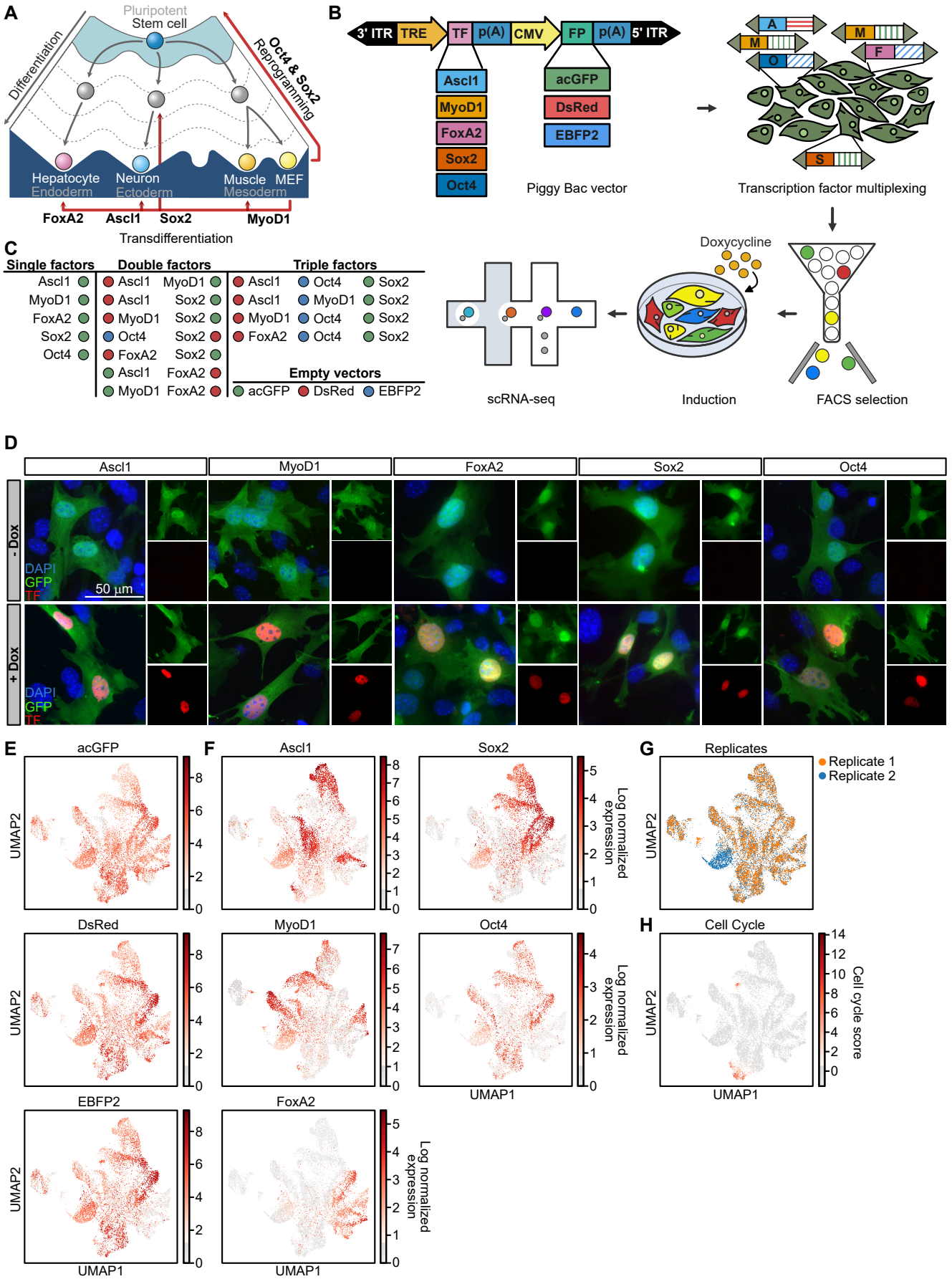

A
Pluripotent
Stem cell
Differentiation
Oct4 & Sox2
Reprogramming
Hepatocyte
Endoderm
Neuron
Ectoderm
Muscle
Mesoderm
MEF
FoxA2
Ascl1
Sox2
MyoD1
Transdifferentiation
B
3' ITR
TRE
TF
p(A)
CMV
FP
p(A)
5' ITR
Ascl1
MyoD1
FoxA2
Sox2
Oct4
acGFP
DsRed
EBFP2
Piggy Bac vector
A
M
O
M
F
S
Transcription factor multiplexing
Doxycycline
scRNA-seq
Induction
FACS selection
C
Single factors
Ascl1
MyoD1
FoxA2
Sox2
Oct4
Double factors
Ascl1 MyoD1
Ascl1 Sox2
MyoD1 Sox2
Oct4 Sox2
FoxA2 Sox2
Ascl1 FoxA2
MyoD1 FoxA2
Triple factors
Ascl1 Oct4 Sox2
Ascl1 MyoD1 Sox2
MyoD1 Oct4 Sox2
FoxA2 Oct4 Sox2
Empty vectors
acGFP DsRed EBFP2
D
Ascl1
MyoD1
FoxA2
Sox2
Oct4
- Dox
+ Dox
DAPI
GFP
TF
50 µm
E
acGFP
DsRed
EBFP2
F
Ascl1
MyoD1
FoxA2
Sox2
Oct4
UMAP1
UMAP2
Log normalized expression
G
Replicates
Replicate 1
Replicate 2
H
Cell Cycle
Cell cycle score

Figure 2

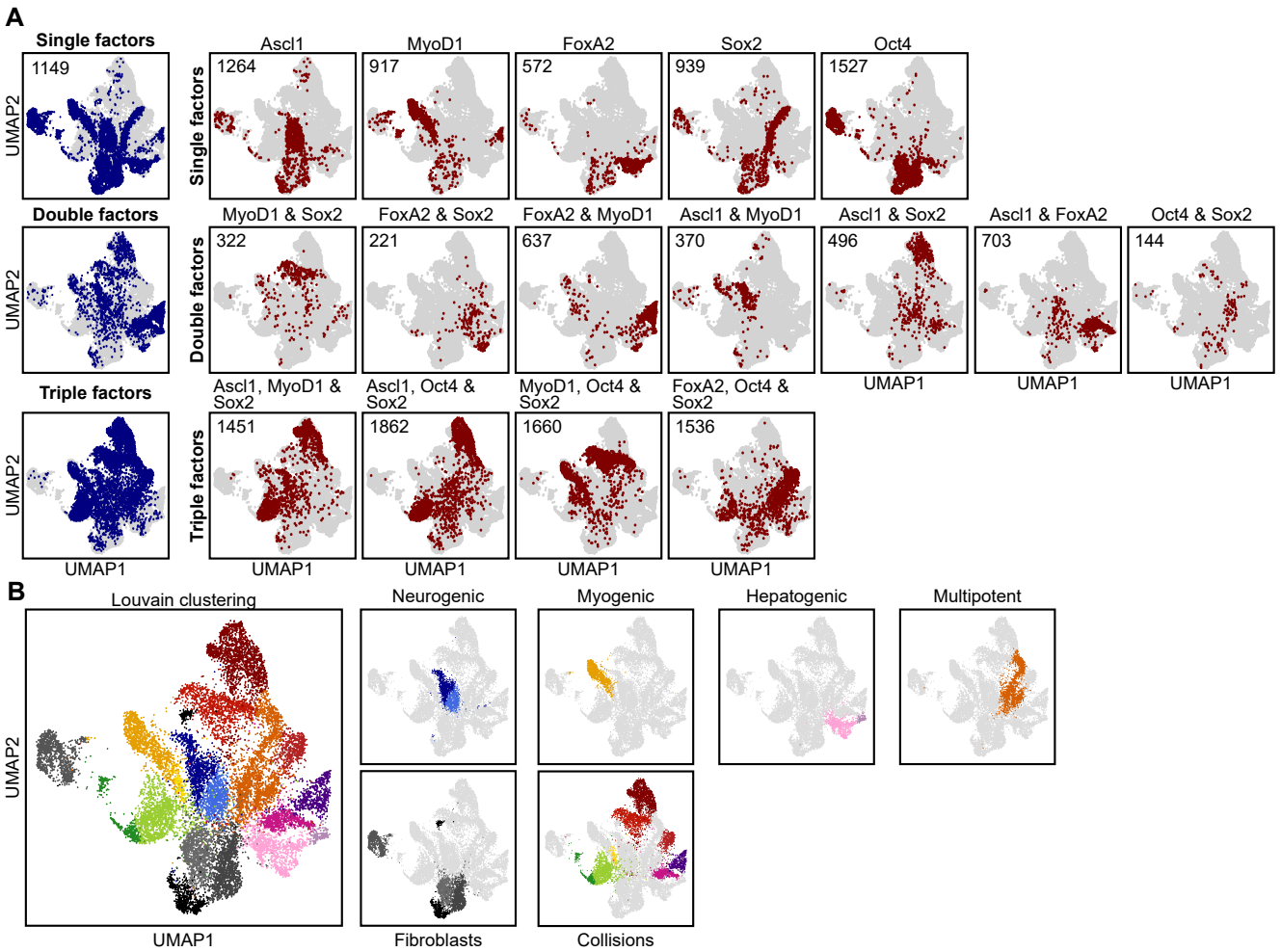

A
Single factors
1149
UMAP2
Ascl1
1264
MyoD1
917
FoxA2
572
Sox2
939
Oct4
1527
Single factors
Double factors
MyoD1 & Sox2
322
FoxA2 & Sox2
221
FoxA2 & MyoD1
637
Ascl1 & MyoD1
370
Ascl1 & Sox2
496
Ascl1 & FoxA2
703
Oct4 & Sox2
144
Double factors
Triple factors
Ascl1, MyoD1 & Sox2
1451
Ascl1, Oct4 & Sox2
1862
MyoD1, Oct4 & Sox2
1660
FoxA2, Oct4 & Sox2
1536
Triple factors
UMAP1
B
Louvain clustering
Neurogenic
Myogenic
Hepatogenic
Multipotent
Fibroblasts
Collisions

Figure 3

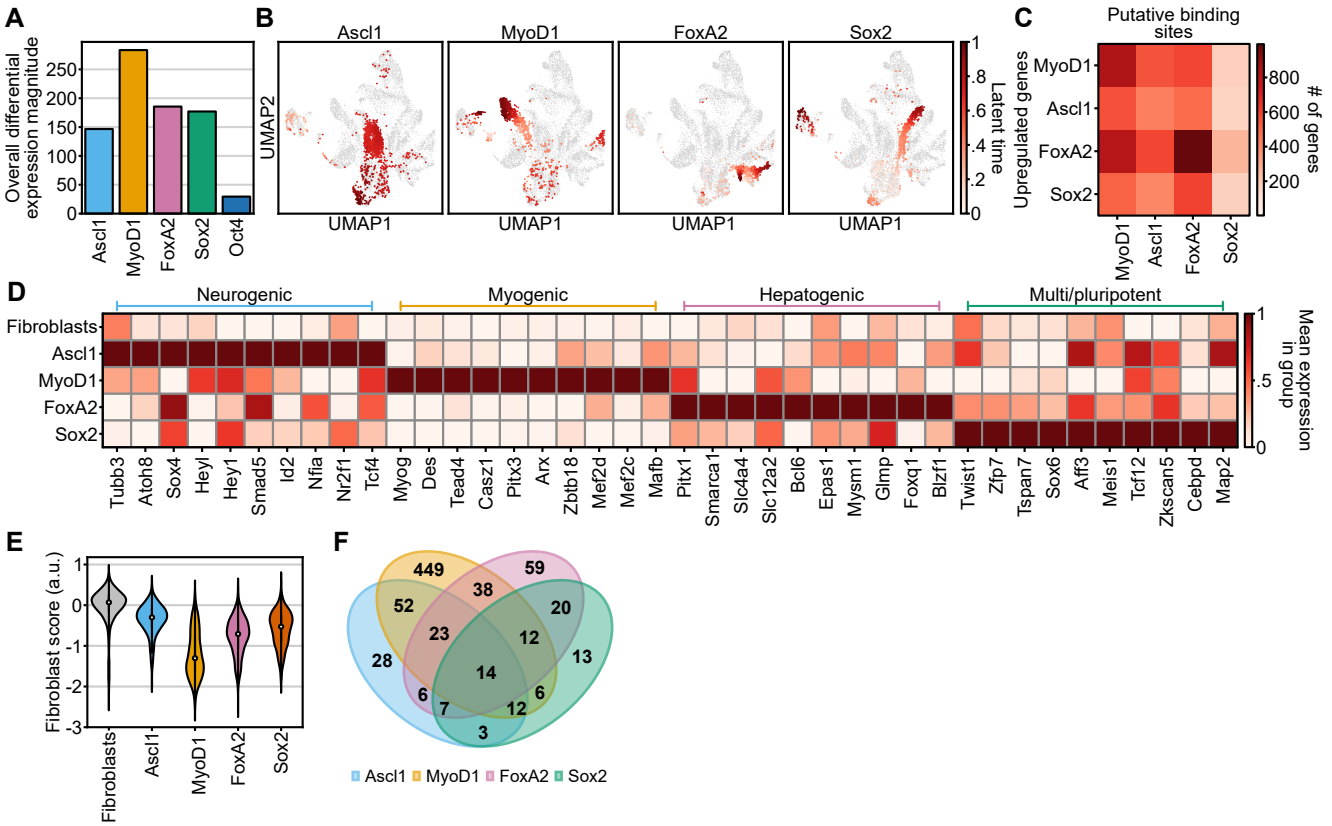

A
Overall differential expression magnitude
Ascl1
MyoD1
FoxA2
Sox2
Oct4
B
Ascl1
MyoD1
FoxA2
Sox2
UMAP2
UMAP1
Latent time
C
Putative binding sites
Upregulated genes
# of genes
D
Neurogenic
Myogenic
Hepatogenic
Multi/pluripotent
Fibroblasts
Mean expression in group
E
Fibroblast score (a.u.)
F
Ascl1 MyoD1 FoxA2 Sox2

Figure 4

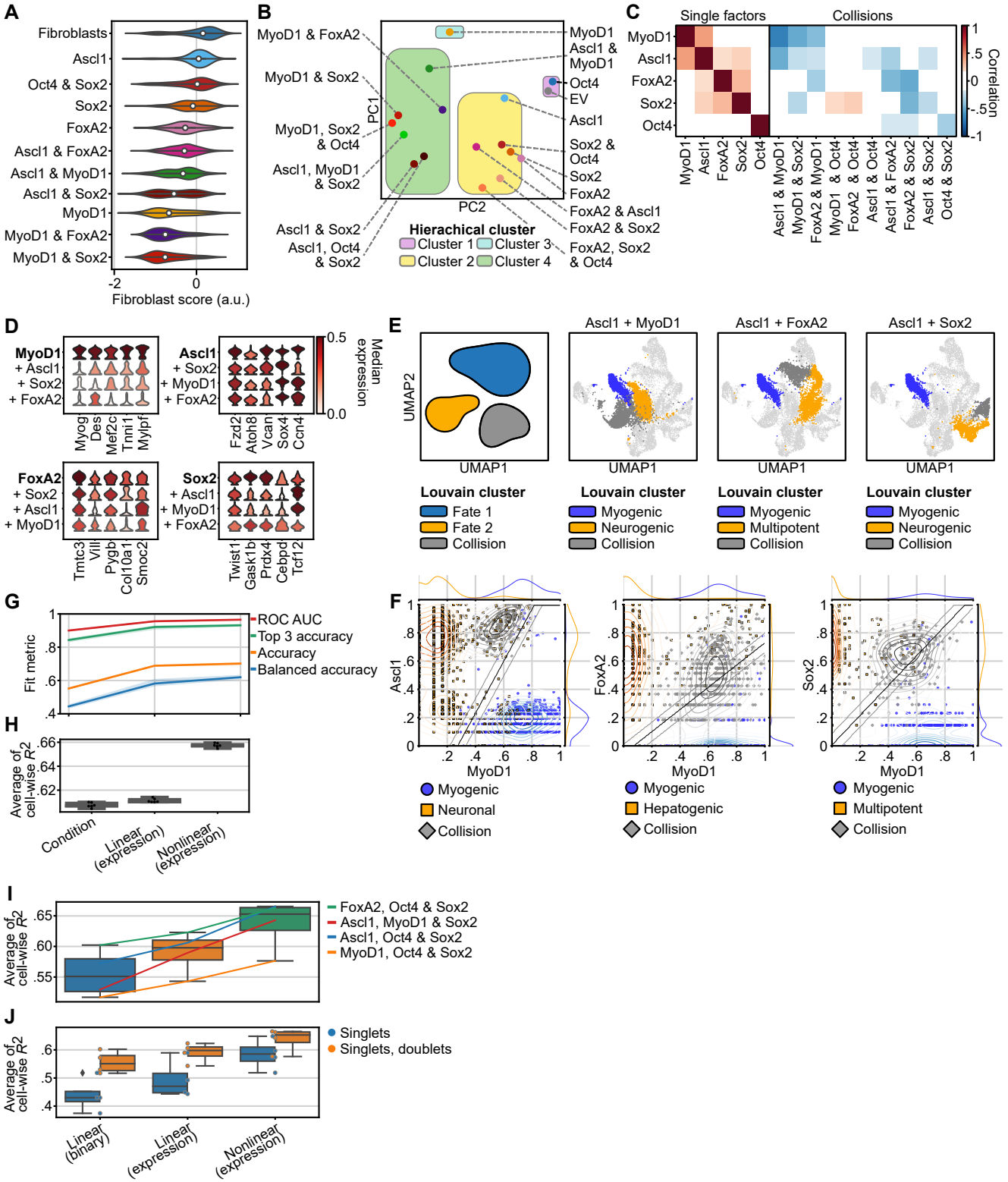

A
Fibroblasts
Ascl1
Oct4 & Sox2
Sox2
FoxA2
Ascl1 & FoxA2
Ascl1 & MyoD1
Ascl1 & Sox2
MyoD1
MyoD1 & FoxA2
MyoD1 & Sox2
-2
0
Fibroblast score (a.u.)
B
MyoD1 & FoxA2
MyoD1 & Sox2
MyoD1, Sox2 & Oct4
Ascl1, MyoD1 & Sox2
Ascl1 & Sox2
Ascl1, Oct4 & Sox2
PC1
PC2
MyoD1
Ascl1 & MyoD1
Oct4
EV
Ascl1
Sox2 & Oct4
Sox2
FoxA2
FoxA2 & Ascl1
FoxA2 & Sox2
FoxA2, Sox2 & Oct4
Hierachical cluster
Cluster 1
Cluster 3
Cluster 2
Cluster 4
C
Single factors
Collisions
MyoD1
Ascl1
FoxA2
Sox2
Oct4
Ascl1 & MyoD1
MyoD1 & Sox2
FoxA2 & MyoD1
MyoD1 & Oct4
FoxA2 & Oct4
Ascl1 & Oct4
Ascl1 & FoxA2
FoxA2 & Sox2
Ascl1 & Sox2
Oct4 & Sox2
1
.5
0
-.5
-1
Correlation
D
MyoD1 + Ascl1 + Sox2 + FoxA2
Myog
Des
Mef2c
Tnni1
Mylpf
Ascl1 + Sox2 + MyoD1 + FoxA2
Fzd2
Atoh8
Vcan
Sox4
Ccn4
0.5
0.0
Median expression
FoxA2 + Sox2 + Ascl1 + MyoD1
Tmtc3
Vill
Pygb
Col10a1
Smoc2
Sox2 + Ascl1 + MyoD1 + FoxA2
Twist1
Gask1b
Prdx4
Cebpd
Tcf12
E
UMAP2
UMAP1
Ascl1 + MyoD1
Ascl1 + FoxA2
Ascl1 + Sox2
Louvain cluster
Fate 1
Fate 2
Collision
Myogenic
Neurogenic
Multipotent
F
Ascl1
FoxA2
Sox2
MyoD1
Myogenic
Neuronal
Hepatogenic
Multipotent
Collision
G
Fit metric
ROC AUC
Top 3 accuracy
Accuracy
Balanced accuracy
H
Average of cell-wise R^2
Condition
Linear (expression)
Nonlinear (expression)
I
FoxA2, Oct4 & Sox2
Ascl1, MyoD1 & Sox2
Ascl1, Oct4 & Sox2
MyoD1, Oct4 & Sox2
J
Singlets
Singlets, doublets
Linear (binary)

Figure 5

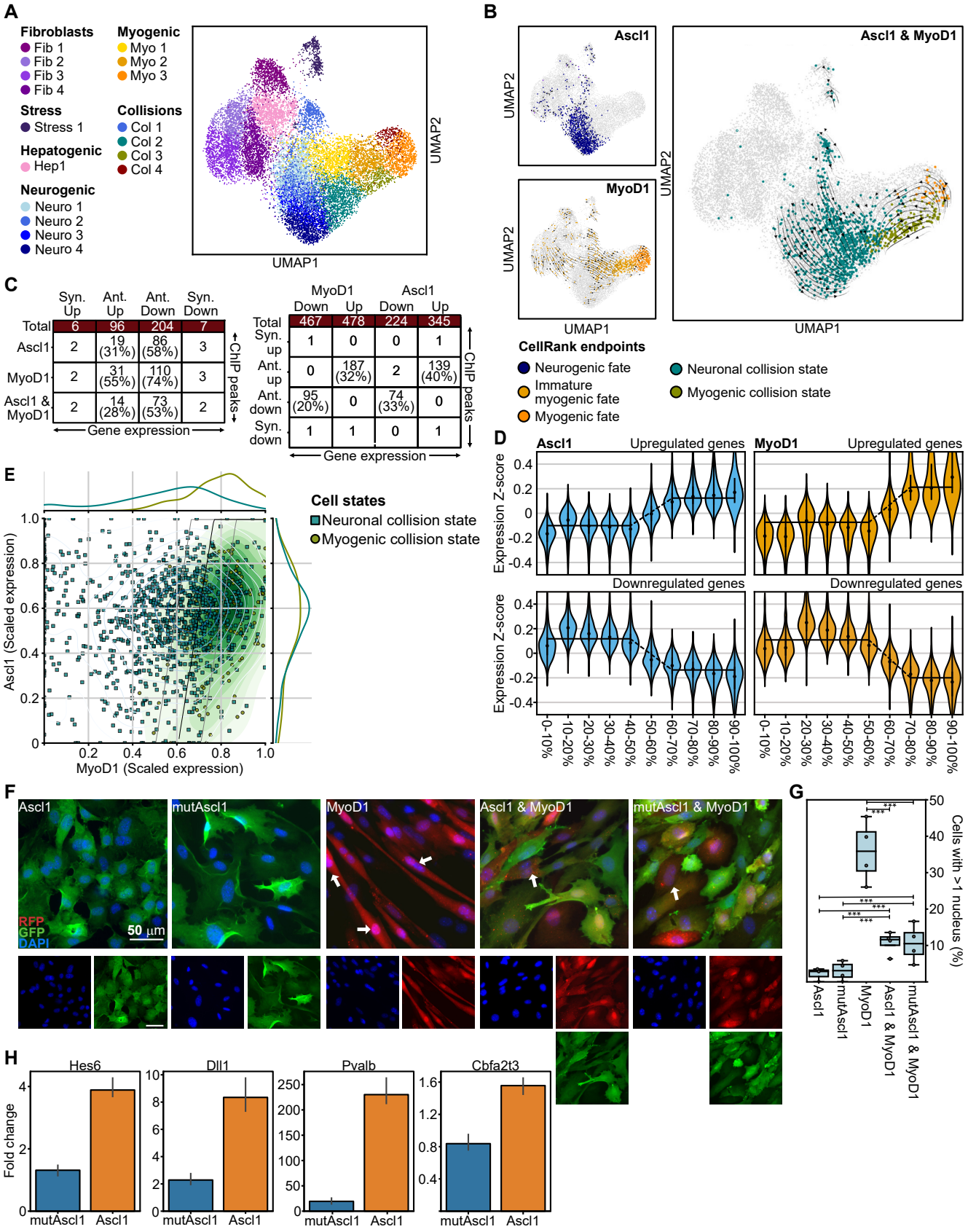

A
Fibroblasts
Fib 1
Fib 2
Fib 3
Fib 4
Stress
Stress 1
Hepatogenic
Hep1
Neurogenic
Neuro 1
Neuro 2
Neuro 3
Neuro 4
Myogenic
Myo 1
Myo 2
Myo 3
Collisions
Col 1
Col 2
Col 3
Col 4
UMAP1
UMAP2
B
Ascl1
Ascl1 & MyoD1
MyoD1
UMAP2
UMAP1
CellRank endpoints
Neurogenic fate
Immature myogenic fate
Myogenic fate
Neuronal collision state
Myogenic collision state
C
Syn. Up
Ant. Up
Ant. Down
Syn. Down
Total 6 96 204 7
Ascl1 2 19 (31%) 86 (58%) 3
MyoD1 2 31 (55%) 110 (74%) 3
Ascl1 & MyoD1 2 14 (28%) 73 (53%) 2
ChIP peaks
Gene expression
MyoD1 Down
MyoD1 Up
Ascl1 Down
Ascl1 Up
Total 467 478 224 345
Syn. up 1 0 0 1
Ant. up 0 187 (32%) 2 139 (40%)
Ant. down 95 (20%) 0 74 (33%) 0
Syn. down 1 1 0 1
ChIP peaks
Gene expression
D
Ascl1
Upregulated genes
MyoD1
Upregulated genes
Downregulated genes
Downregulated genes
Expression Z-score
0-10%
10-20%
20-30%
30-40%
40-50%
50-60%
60-70%
70-80%
80-90%
90-100%
E
Cell states
Neuronal collision state
Myogenic collision state
Ascl1 (Scaled expression)
MyoD1 (Scaled expression)
F
Ascl1
mutAscl1
MyoD1
Ascl1 & MyoD1
mutAscl1 & MyoD1
RFP
GFP
DAPI
50 µm
G
Cells with >1 nucleus (%)
Ascl1
mutAscl1
MyoD1
Ascl1 & MyoD1
mutAscl1 & MyoD1
H
Hes6
Dll1
Pvalb
Cbfa2t3
Fold change
mutAscl1
Ascl1

Figure 6

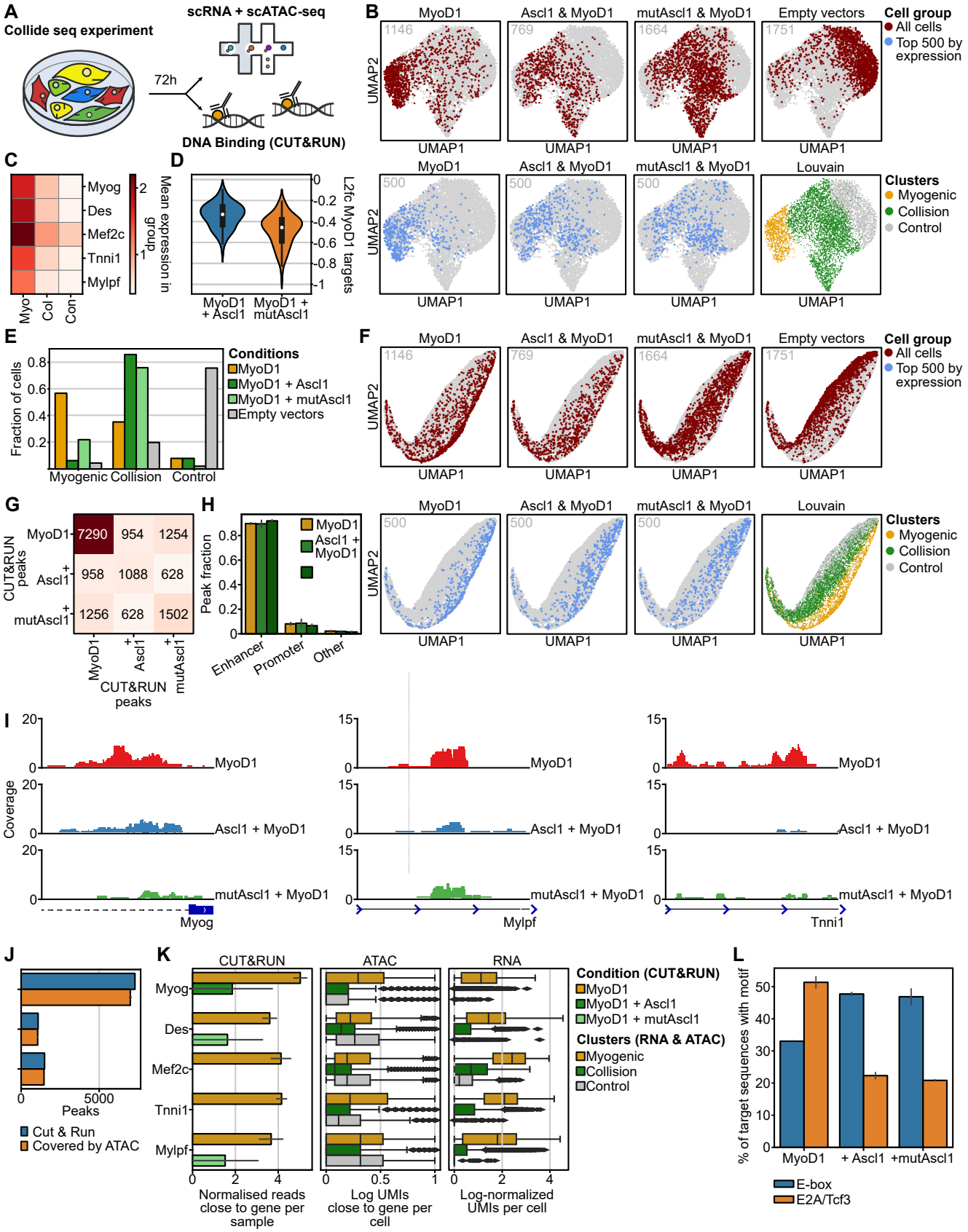
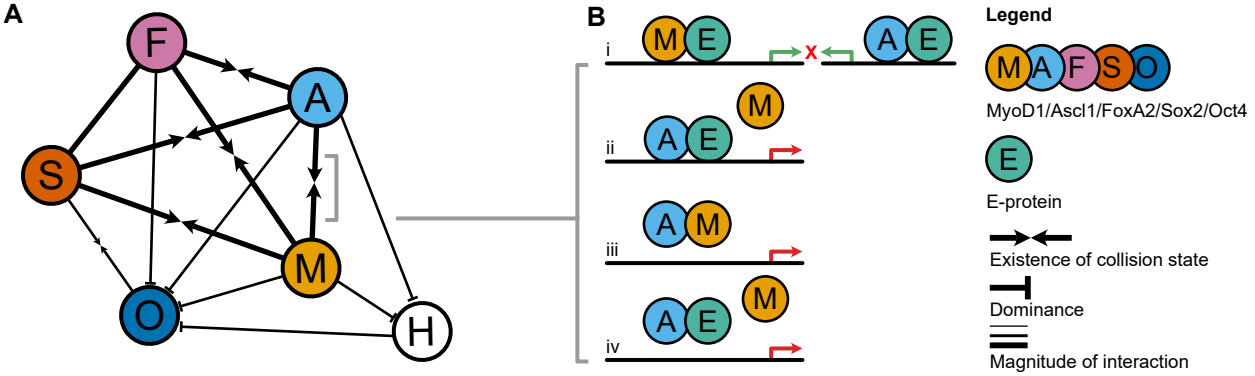
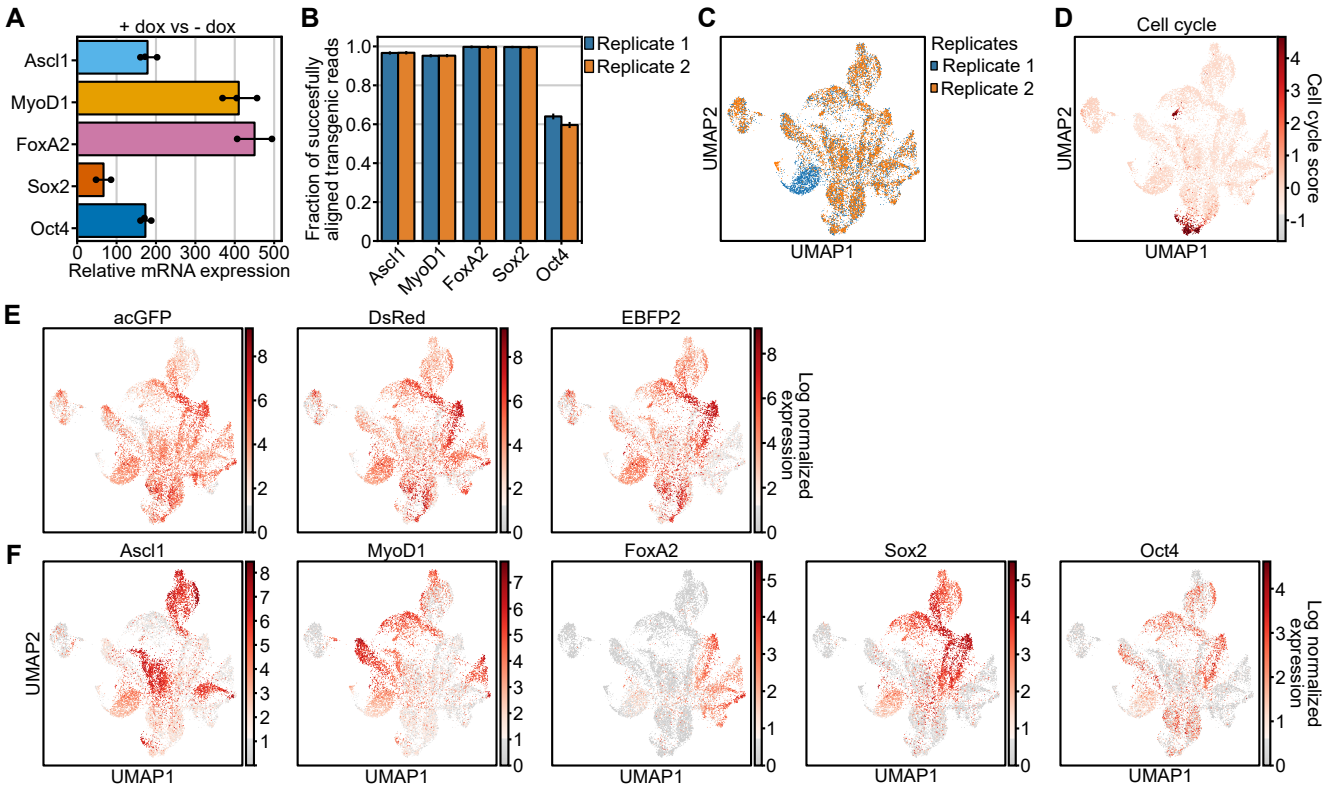

A
Collide seq experiment
scRNA + scATAC-seq
72h
DNA Binding (CUT&RUN)
B
MyoD1
Ascl1 & MyoD1
mutAscl1 & MyoD1
Empty vectors
Louvain
UMAP1
UMAP2
Cell group
All cells
Top 500 by expression
Clusters
Myogenic
Collision
Control
C
Myog
Des
Mef2c
Tnni1
Mylpf
Mean expression in group
D
L2fc MyoD1 targets
MyoD1 + Ascl1
MyoD1 + mutAscl1
E
Fraction of cells
Conditions
MyoD1
MyoD1 + Ascl1
MyoD1 + mutAscl1
Empty vectors
Myogenic
Collision
Control
F
G
CUT&RUN peaks
7290
954
1254
958
1088
628
1256
628
1502
H
Peak fraction
Enhancer
Promoter
Other
I
Coverage
Myog
Mylpf
Tnni1
J
Peaks
Cut & Run
Covered by ATAC
K
CUT&RUN
ATAC
RNA
Normalised reads close to gene per sample
Log UMIs close to gene per cell
Log-normalized UMIs per cell
Condition (CUT&RUN)
Clusters (RNA & ATAC)
L
% of target sequences with motif
E-box
E2A/Tcf3

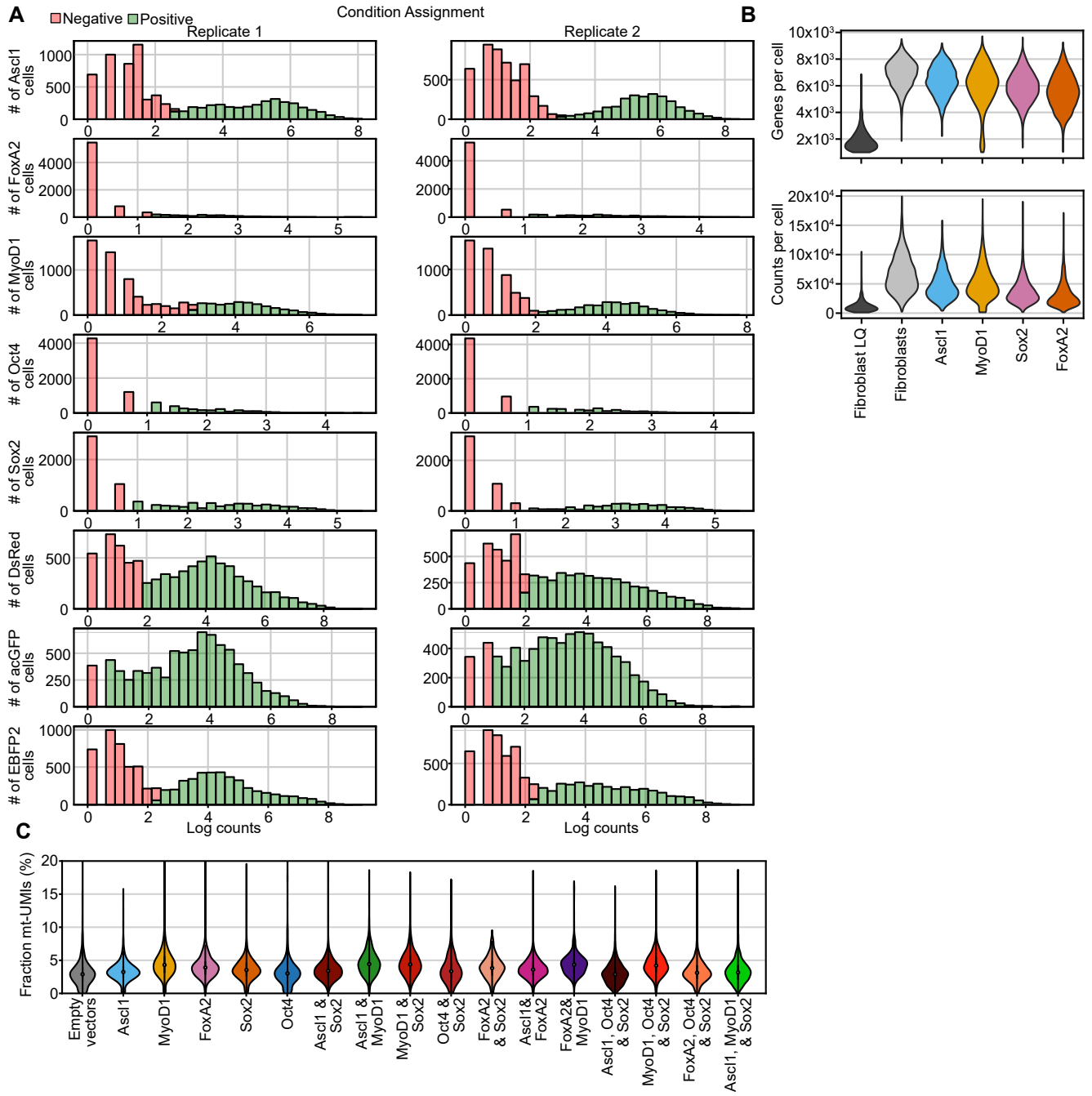
Figure 7


A
F
A
S
M
O
H
B
i
M
E
x
A
E
ii
M
A
E
iii
A
M
iv
M
A
E
Legend
M A F S O
MyoD1/Ascl1/FoxA2/Sox2/Oct4
E
E-protein
Existence of collision state
Dominance
Magnitude of interaction

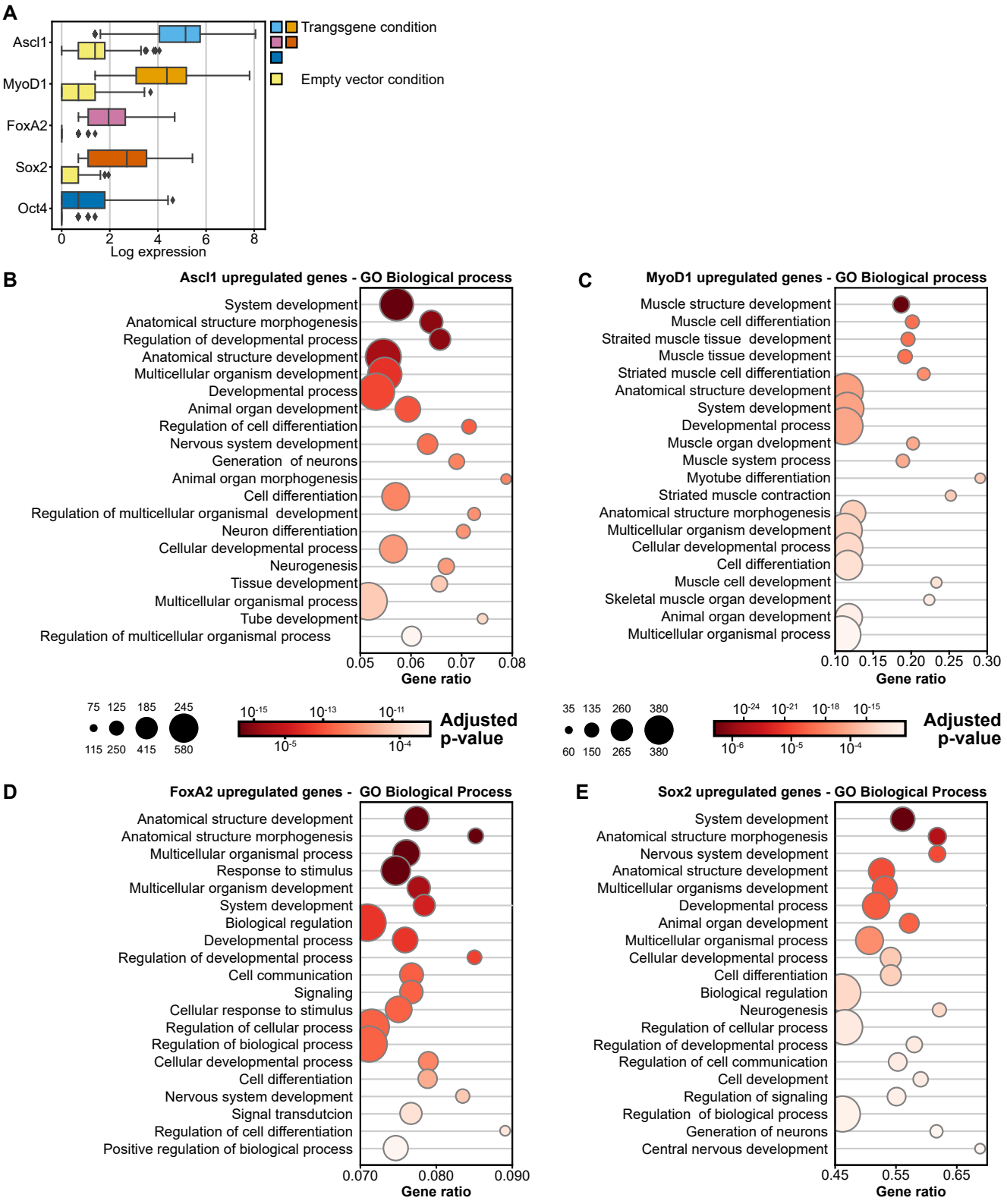
Supplementary Figure 1



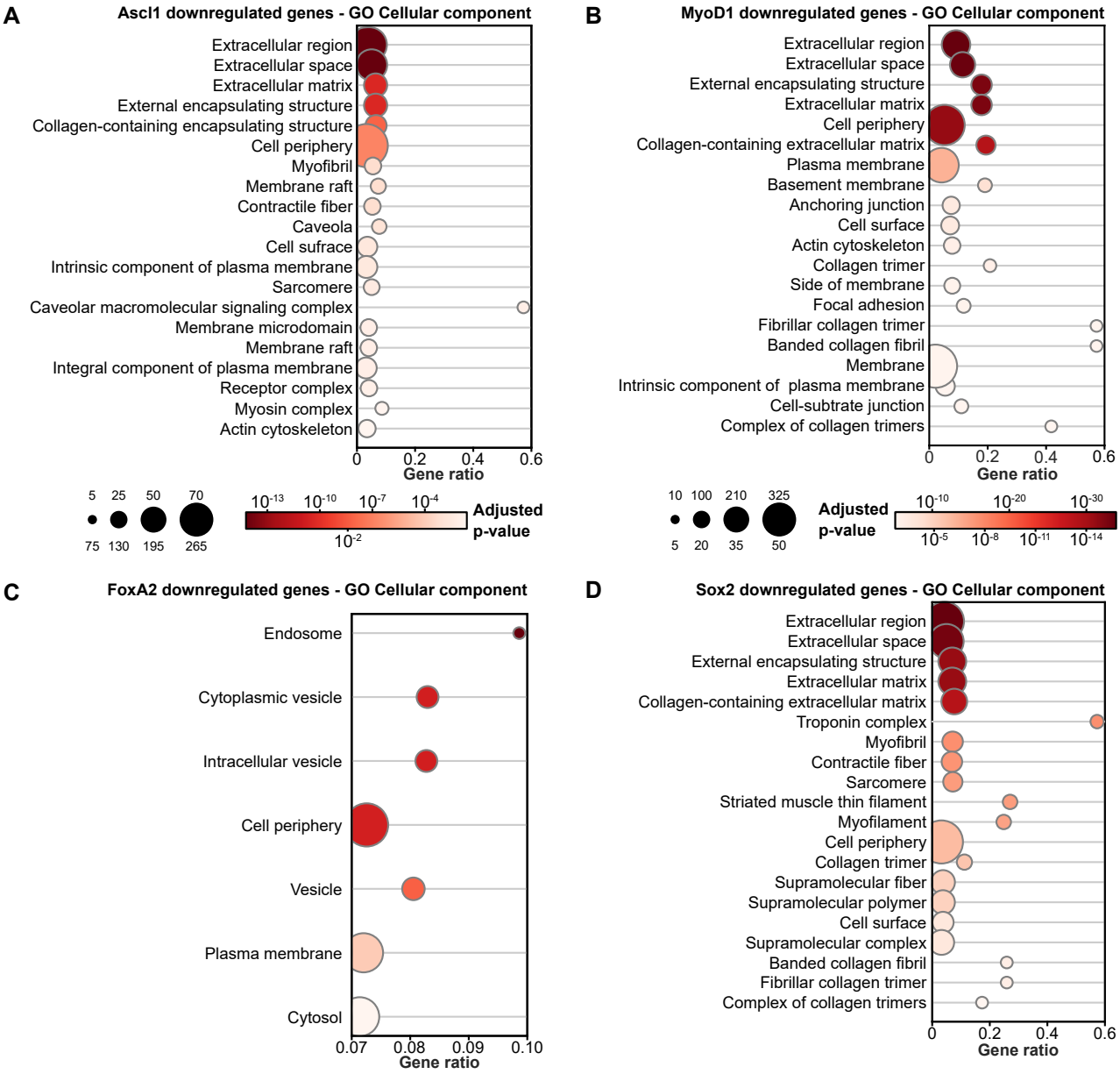
Supplementary Figure 2



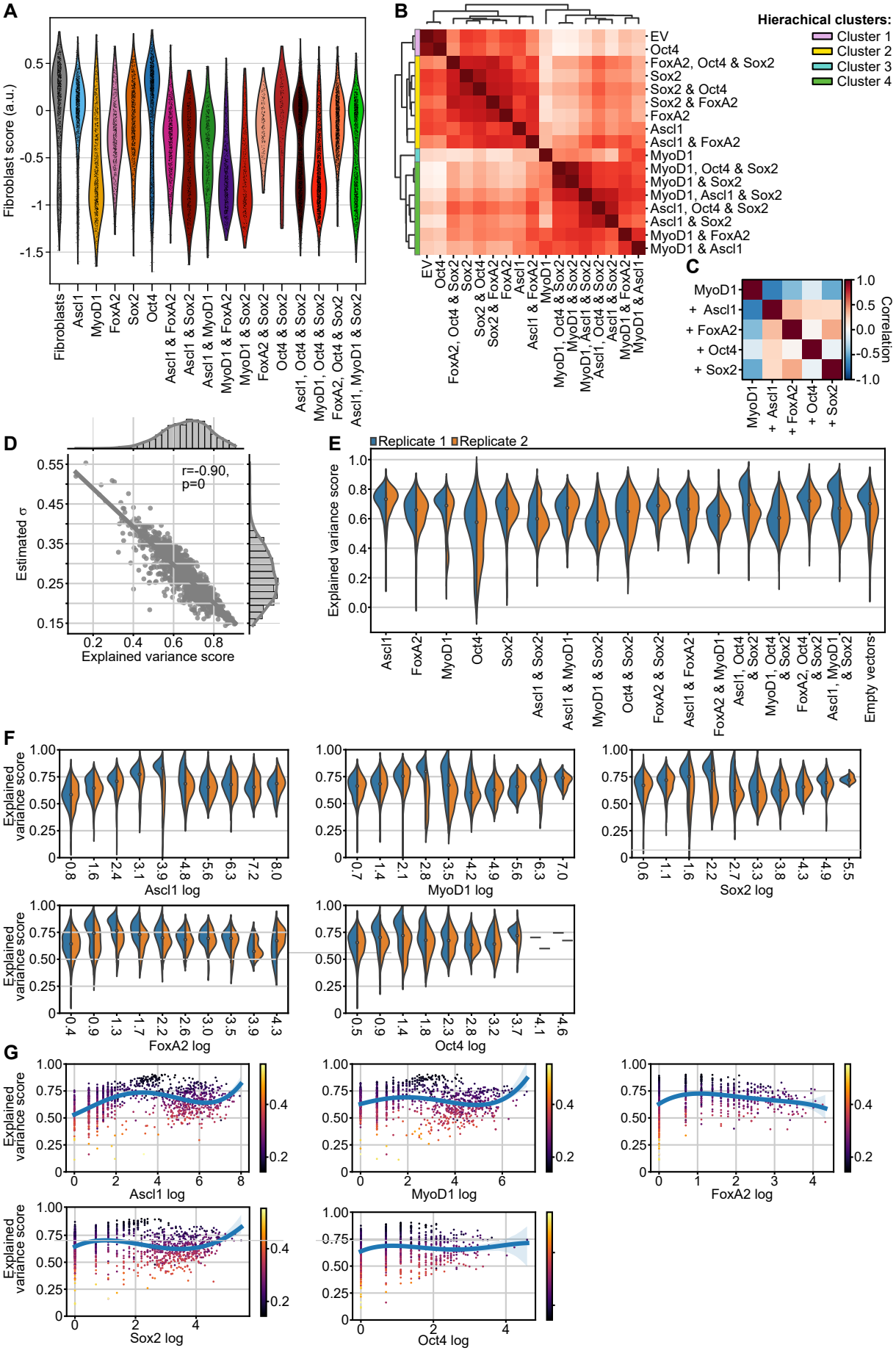
Supplementary Figure 3



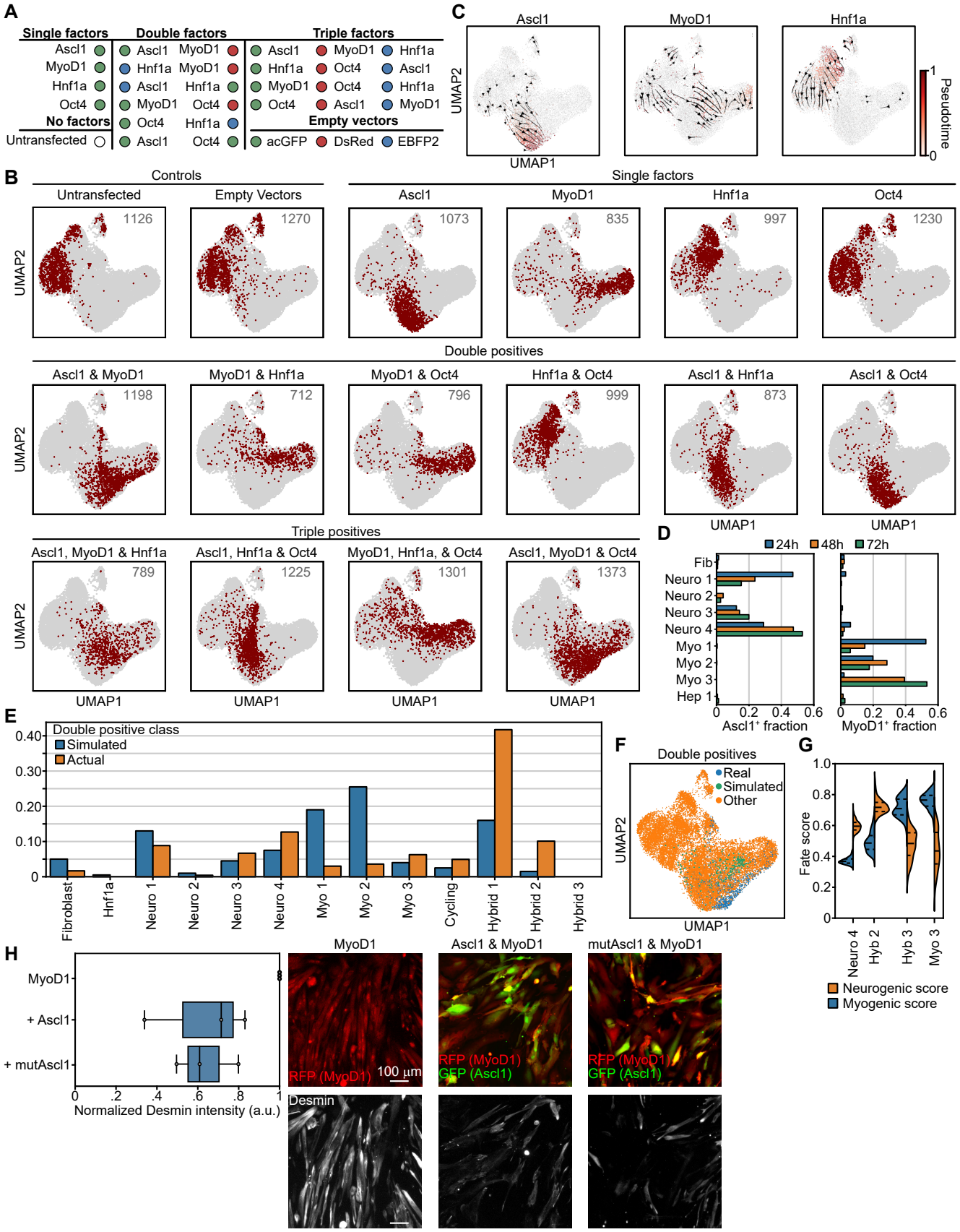
Supplementary Figure 4



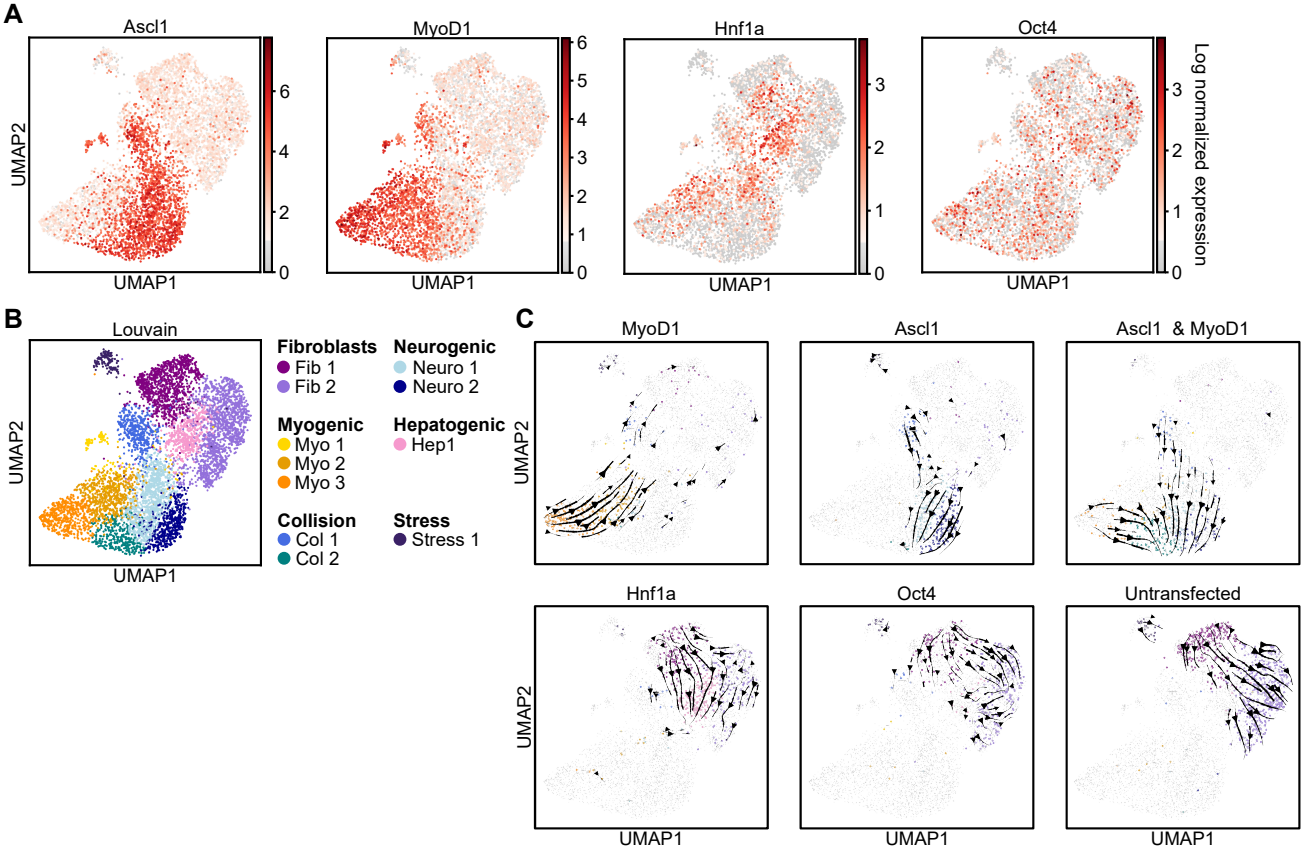
Supplementary Figure 5



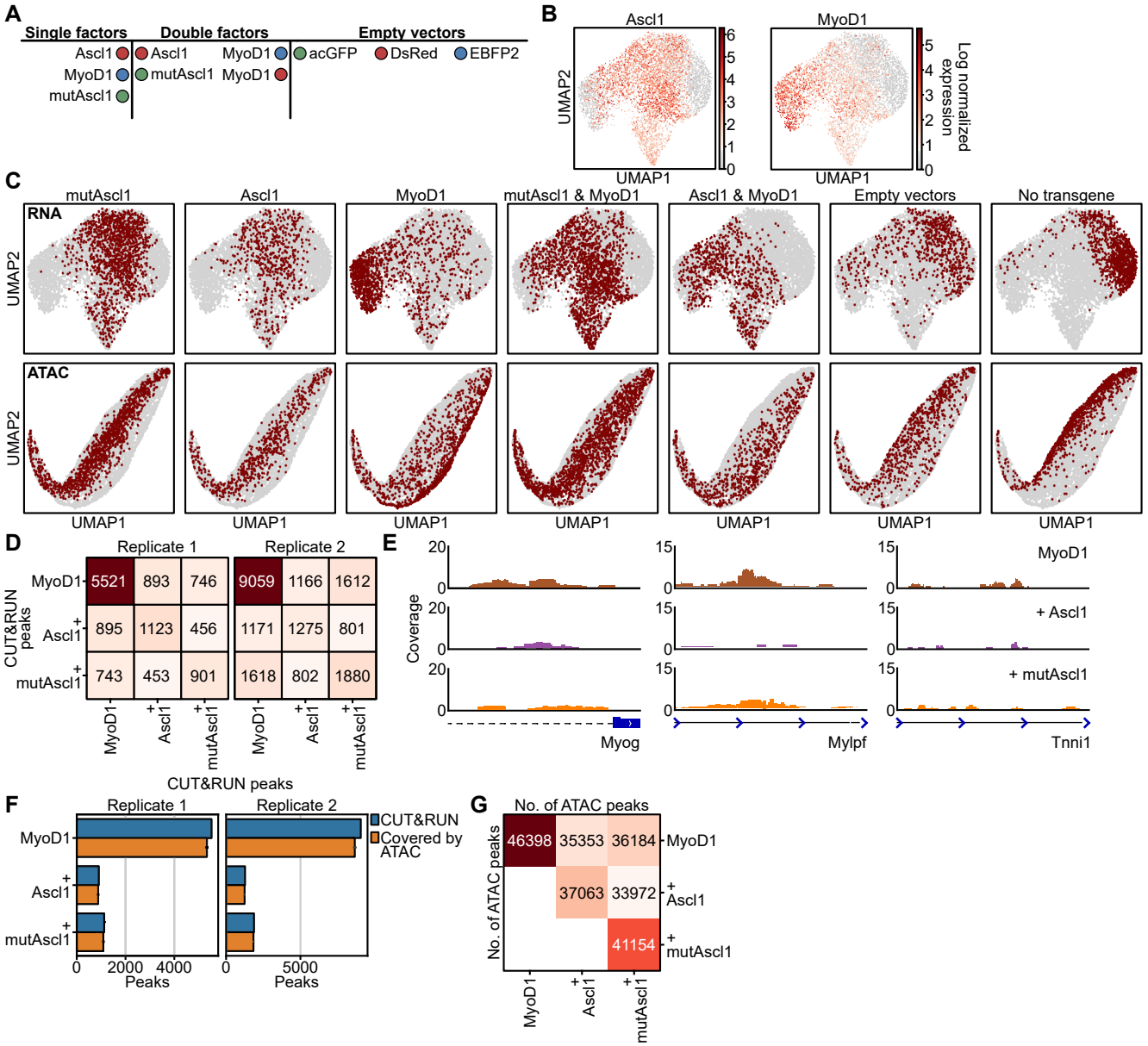
Supplementary Figure 6



Supplementary Figure 7



Supplementary Figure 8



RESULTS II

Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2

J.Kempf,* K.Knelles*, B.A.Hersbach*, D.Petrik, T.Riedemann, V.Bednarova, A.Janjic, T.Simon-Ebert, W.Enard, P.Smialowski, M.Götz*, G.Masserdotti*

*: These authors contributed equally

Author contributions

G.M. conceived and designed the project. J.K. and K.K. performed the isolation, culturing, and most of the reprogramming experiments. J.K. and K.K. evaluated the reprogramming efficiency at 8DPI; J.K. performed and quantified the experiments at day 21. K.K. prepared the cells for FACS analysis and extracted RNA for bulk RNA-seq analysis. **B.A.H.** prepared primary cultures of astrocytes from cortex GM, collected single cells, processed the samples with SmartSeq2, and prepared the libraries for sequencing; D.P. collected single iNs for patch-seq, processed them via Smart-Seq2 protocol, and performed morphometric analysis. V.B performed patch-seq experiments, analyzed the electrophysiological data, prepared the samples via Smart-Seq2, and performed initial analysis of RNA-seq data. A.J. prepared bulk-adapted mcSCRB-seq libraries; T.S.-E. contributed to establish MACS protocol for SC, trained J.K. and K.K., and performed immunocytochemistry; T.R. performed whole-cell patch-clamp experiments and analyzed the electrophysiological properties of iNs. W.E. provided reagents; G.M. performed FACS and analyzed data from bulk and single-cell RNA-seq; P.S. compared single-cell RNA with publicly available data and analyzed single-cell RNA-seq with G.M.; M.G. contributed to the conceptual organization of the manuscript and the patch-seq data and financed the work. G.M. and M.G. wrote the manuscript, and **all authors contributed corrections and comments.**

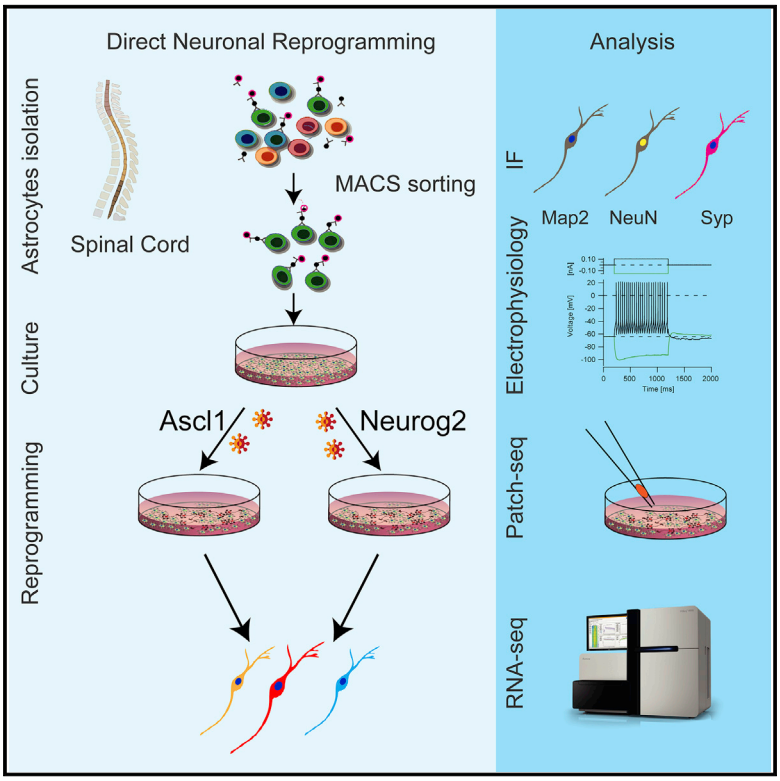
Permissions

Cell Reports

Article

Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2

Graphical abstract



Highlights

- Ascl1 and Neurog2 induce initially distinct transcriptomes in spinal cord astrocytes
- Neurons induced by Ascl1 or Neurog2 converge to a V2 interneuron-like state
- Patch-seq shows functional and transcriptional heterogeneity with low correlation
- Developmentally established patterning genes are maintained in astrocytes *in vitro*

Authors

J. Kempf, K. Knelles, B.A. Hersbach, ..., P. Smialowski, M. Götz, G. Masserdotti

Correspondence

magdalena.goetz@helmholtz-muenchen.de (M.G.), giacomo.masserdotti@helmholtz-muenchen.de (G.M.)

In brief

Kempf et al. show that Ascl1 and Neurog2 elicit initially divergent transcriptional programs in spinal cord astrocytes converging later to a V2 interneuron-like state, according to their developmental role and patterning cues in astrocytes. Patch-seq analysis reveals the heterogeneity of fate conversion with little correlation between transcriptome and electrophysiological properties.

Kempf et al., 2021, Cell Reports *36*, 109409
July 20, 2021
https://doi.org/10.1016/j.celrep.2021.109409

Cell Reports

CellPress
OPEN ACCESS

Article

Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2

J. Kempf,[1,8] K. Knelles,[1,8] B.A. Hersbach,[1,2,3,8] D. Petrik,[1,2,4] T. Riedemann,[1] V. Bednarova,[1] A. Janjic,[6] T. Simon-Ebert,[1] W. Enard,[5] P. Smialowski,[1,2,5] M. Götz,[1,2,7,8,9,*] and G. Masserdotti[1,2,8,9,*]

[1]Biomedical Center Munich, Physiological Genomics, LMU Munich, Planegg-Martinsried 82152, Germany
[2]Institute for Stem Cell Research, Helmholtz Center Munich, Neuherberg 85764, Germany
[3]Graduate School of Systemic Neurosciences, LMU Munich, Planegg-Martinsried 82152, Germany
[4]School of Biosciences, The Sir Martin Evans Building, Cardiff University, CF10 3AX Cardiff, UK
[5]Biomedical Center Munich, Bioinformatic Core Facility, LMU Munich, Planegg-Martinsried 82152, Germany
[6]Anthropology and Human Genomics, Faculty of Biology, LMU Munich, Planegg-Martinsried 82152, Germany
[7]Excellence Cluster of Systems Neurology (SYNERGY), Munich, Germany
[8]These authors contributed equally
[9]Lead contact
*Correspondence: magdalena.goetz@helmholtz-muenchen.de (M.G.), giacomo.masserdotti@helmholtz-muenchen.de (G.M.)
https://doi.org/10.1016/j.celrep.2021.109409

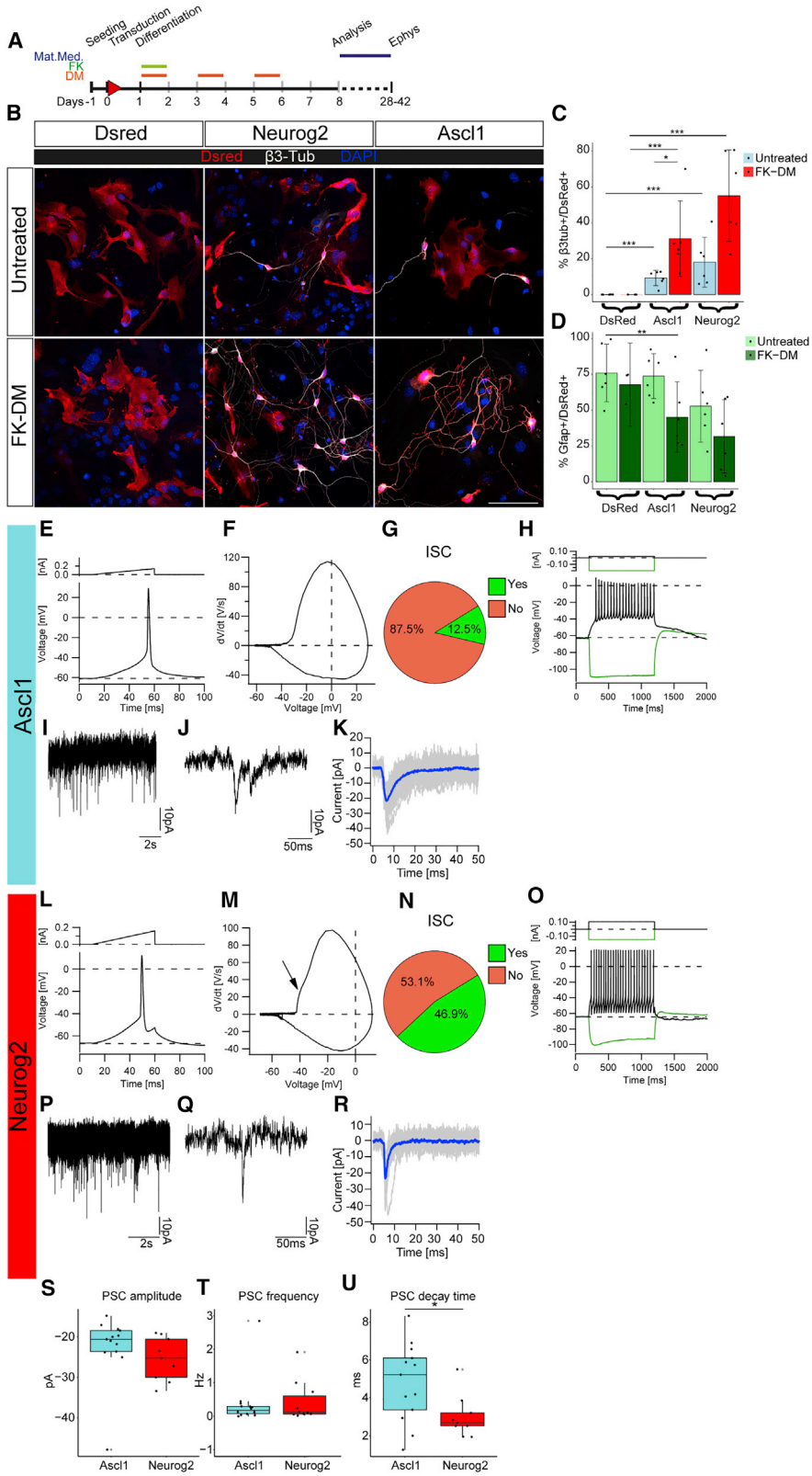
SUMMARY

Astrocytes are a viable source for generating new neurons via direct conversion. However, little is known about the neurogenic cascades triggered in astrocytes from different regions of the CNS. Here, we examine the transcriptome induced by the proneural factors Ascl1 and Neurog2 in spinal cord-derived astrocytes *in vitro*. Each factor initially elicits different neurogenic programs that later converge to a V2 interneuron-like state. Intriguingly, patch sequencing (patch-seq) shows no overall correlation between functional properties and the transcriptome of the heterogenous induced neurons, except for K-channels. For example, some neurons with fully mature electrophysiological properties still express astrocyte genes, thus calling for careful molecular and functional analysis. Comparing the transcriptomes of spinal cord- and cerebral-cortex-derived astrocytes reveals profound differences, including developmental patterning cues maintained *in vitro*. These relate to the distinct neuronal identity elicited by Ascl1 and Neurog2 reflecting their developmental functions in subtype specification of the respective CNS region.

INTRODUCTION

Direct reprogramming of local glial cells into neurons in the central nervous system (CNS) has become a promising approach for neuronal replacement in disease (Barker et al., 2018). Pioneered by *in vitro* conversion of astrocytes from the cerebral cortex (Berninger et al., 2007; Heinrich et al., 2010; Heins et al., 2002), this approach is often based on the expression of proneural factors (e.g., Ascl1 and Neurog2), master regulators and pioneer factors in the conversion process both *in vitro* (Berninger et al., 2007; Masserdotti et al., 2015; Smith et al., 2016; Vierbuchen et al., 2010; Wapinski et al., 2013) and *in vivo* (Guo et al., 2014; Liu et al., 2015; Mattugini et al., 2019; Rivetti di Val Cervo et al., 2017; Torper et al., 2013). Cortical astrocyte cultures allowed identifying major reprogramming hurdles (Gascón et al., 2016; Russo et al., 2020), whose manipulation *in vivo* led to improving the reprogramming efficiency from 10% to over 90% (Gascón et al., 2016).

However, the use of neuronal reprogramming for therapy requires the generation of adequate neuronal subtypes specific to different CNS regions. For instance, the induction of most of the spinal cord (SC) neuronal diversity has not yet been achieved (Su et al., 2014b), besides motoneuron generation from fibroblasts (Abernathy et al., 2017; Church et al., 2021; Meyer et al., 2014; Son et al., 2011) and neurons with different neurotransmitter identities from NG2 glia *in vivo* (Tai et al., 2021). Ascl1 and Neurog2 are involved in generating specific interneuron subtypes in the developing SC (Lu et al., 2015; Misra et al., 2014): Ascl1 specifies excitatory neurons (Borromeo et al., 2014; Mizuguchi et al., 2006), and Neurog2 is a downstream effector of Ptf1a, instructing GABAergic interneurons (Henke et al., 2009) or motorneurons (Lee et al., 2020). In the ventral SC, both Ascl1 and Neurog2 specify a V2 interneuron identity (Parras et al., 2002), specifically a GABAergic V2b interneuron fate (Misra et al., 2014). In forebrain development, Ascl1 instructs inhibitory GABAergic neurons and Neurog1/2 glutamatergic neurons (Kovach et al., 2013; Parras et al., 2002), which is reflected in the neuronal subtypes generated by these proneural genes from cortical astrocytes (Berninger et al., 2007; Heinrich et al., 2010; Masserdotti et al., 2015). Recently, differences in the



(legend on next page)

reprogramming efficiency of astrocytes obtained from various CNS regions have been reported (Chouchane et al., 2017; Hu et al., 2019; Rao et al., 2021). However, the mechanisms underlying direct neuronal reprogramming of region-specific astrocytes are largely unknown. Here, we investigated the neurogenic programs induced by Ascl1 and Neurog2 in SC-derived astrocytes, the neuronal identity of reprogrammed neurons, and how the regional specification of SC astrocytes may influence the fate of neurons reprogrammed by Ascl1 or Neurog2.

RESULTS

Direct conversion of SC-derived astrocytes into functional neurons

Protocols used for culturing cortical astrocytes (Heinrich et al., 2011) (Figure S1A) were not successful to enrich for SC astrocytes from postnatal day (P) 2–3 mice, as glial fibrillary acidic protein (Gfap)-positive astrocytes were below 50% (Figures S1B and S1C; mean = 41.7%, confidence interval [CI] = 20.6). To enrich for astrocytes, we isolated ACSA-2+ cells (Kantzer et al., 2017) via magnetic-activated cell sorting (MACS). Indeed, the majority of sorted cells were now astrocytes after 8 days in culture (Gfap+, mean = 79.4%, CI = 9.8; Sox9+: mean = 85.9%, CI = 7.6; Figures S1D and S1E) with a low abundance of oligodendrocyte progenitors (Olig2+, mean = 6%, CI = 11.6; expression data in Figure S1H) or microglia (Iba1+, mean = 1%, CI = 1) and no detectable neuroblasts (Dcx+) or neurons (β3-tubulin+) (Figure S1E). Thus, ACSA2-MACS yielded highly enriched cultures of astrocytes to explore astrocyte-to-neuron reprogramming.

SC-derived astrocytes were transduced with retrovirus expressing either Ascl1 or Neurog2 under the CAG promoter (Heinrich et al., 2010, 2011), and, 8 days post-infection (DPI), the proportion of β3-tubulin immunoreactive cells with neuronal morphology among transduced cells was quantified (Gascón et al., 2016; Masserdotti et al., 2015) (Figure 1A). A relatively low proportion of neuronal cells was observed (Neurog2, mean = 9.21%, CI = 6.23; Ascl1, mean = 18.07%, CI = 21.11; Figures 1B and 1C; Gfap+/DsRed+ in Figure 1D), in line with previous reports (Hu et al., 2019) and considerably lower than the efficiency observed in cerebral cortex gray matter (GM)-derived astrocytes, using the same factors and culture conditions (~60% for Neurog2, ~40% for Ascl1) (Heinrich et al., 2010; Hu et al., 2019). Treatment with small molecules (forskolin and dorsomorphin) (Gascón et al., 2016; Liu et al., 2013; Smith et al., 2016) remarkably enhanced the conversion efficiency of SC astrocytes by about 3 times (Neurog2, mean = 31.20%, CI = 19.30; Ascl1, mean = 55.15%, CI = 45.11; Figures 1B and 1C), suggesting the presence of similar reprogramming hurdles as observed in other cells (Gascón et al., 2016; Liu et al., 2013). Notably, few oligodendrocyte progenitors were also detected in these cultures (Figures S1F and S1G).

Electrophysiology performed at 4–6 weeks after infection (28–42 DPI) showed that most reprogrammed neuronal cells generated an action potential (75% and 84.4% of Ascl1- and Neurog2-iNeurons [iNs], respectively; Figures S1I and S1J) and exhibited similar passive and active membrane properties (Figures S1K–S1P); the action potential duration was significantly higher in Ascl1- than in Neurog2-reprogrammed neurons (Figure S1M), suggesting a lower density of Na^+-channels in Ascl1-converted neurons. Single evoked action potential (Figures 1E and 1L) showed an initial spike component (ISC, so-called prepotential; Crochet et al., 2004; Golding and Spruston, 1998; Figures 1F, 1G, 1M, and 1N), while suprathreshold depolarizing current pulses induced repetitive spikes in both Ascl1- and Neurog2-converted neurons (Figures 1H, 1O, S1Q, and S1R). Importantly, both Ascl1- and Neurog2-reprogrammed neurons showed post-synaptic currents (PSC; Figures 1I–1K and 1P–1R), with similar amplitude and frequency (Figures 1S and 1T) but different decay times (Figure 1U). Thus, Ascl1 and Neurog2 reprogrammed SC-derived astrocytes into functional iNs (Yang et al., 2011).

As the electrophysiological properties suggested some differences between the Ascl1- and Neurog2-iNs, we explored the reprogramming at early and late time points at the molecular level.

Transcriptional programs induced by Ascl1 and Neurog2 at early stages in SC astrocytes

To analyze the programs induced at early stages, we timed the induction of target genes using our previously established approach with the hydroxytamoxifen (OHT)-inducible forms of Ascl1 and Neurog2 (Masserdotti et al., 2015). After 24 h of OHT treatment, Ascl1ERT2 and Neurog2ERT2 localized to the nucleus (Figures S2A–S2C), and their protein level was similar

Figure 1. Direct conversion of SC-derived astrocytes into functional neurons

(A) Scheme of the experimental design.

(B) Micrographs depicting cells transduced with control (DsRed), Ascl1-, and Neurog2-encoding virus and converted into neuronal cells. Forskolin (FK) and dorsomorphin (DM) treatment is shown (lower panels). Scale bar: 100 μm.

(C and D) Barplot of %β3-tub+/DsRed+ cells with neuronal morphology (C) and %Gfap+/DsRed+ (D) at 8DPI with mean and CI (95% confidence for n = 6). Linear regression model performed. n= 6 independent experiments; *p < 0.05; **p < 0.001; ***p < 0.0001.

(E, F, L, and M) Examples of a single action potential (E and L) and corresponding phaseplane plots of the action potential in Ascl1- (F) or Neurog2- (M) iNs. Arrow (M) indicates the initial spike component (ISC).

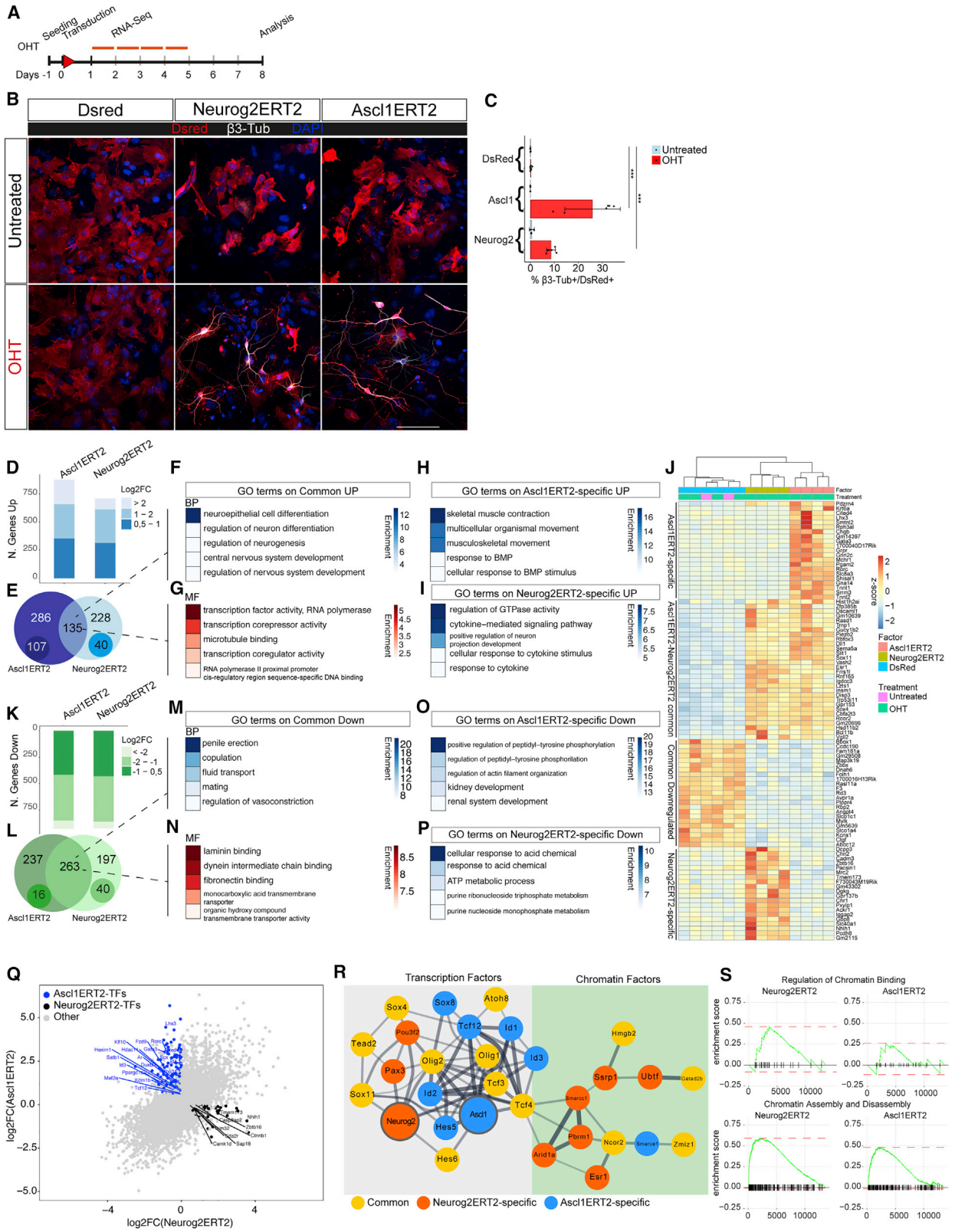
(G and N) Pie charts depicting the percent of cells with ISC among iNs (3/24 and 15/32 cells recorded, respectively). n = 3 independent experiments.

(H and O) Example of an evoked train of action potentials (Ascl1-iNs, H; Neurog2-iNs, O). The green line shows a hyperpolarizing voltage response upon injection of hyperpolarizing current steps (1 s, 150 pA), characterized by a decline of amplitude during ongoing current injection (H) or without such decline (O). Injection of suprathreshold depolarizing current pulses (1 s, 105 pA, upper traces, black) evoked a continuous pattern train of action potentials (H and O), followed by a marked depolarization and long-lasting after-hyperpolarization.

(I, J, P, and Q) Examples of spontaneous synaptic activity in cultures transduced with Ascl1 (I, magnification in J) or Neurog2 (P, magnification in Q).

(K and R) Graphs of the averaged PSC from the cells in (J) and (R), respectively.

(S–U) Boxplots showing PSC amplitude (S), frequency (T), and decay time (U) in Ascl1 and Neurog2 iNs. Each dot represents a cell. *p < 0.05.



(legend on next page)

(Figure S2D). Following 4 days of OHT treatment, both factors converted SC astrocytes into neuronal cells with similar efficiency as the constitutively expressed factors at 8 DPI (Ascl1ERT2, β3-tubulin+/DsRed+ cells, mean = 25.92%, CI = 20.35; Neurog2ERT2, mean = 8.64%, CI = 3.38; no small molecules; Figures 2A–2C). For RNA sequencing (RNA-seq) (Bagnoli et al., 2018), transduced cells were selected via fluorescence-activated cell sorting (FACS; Figure S2E) 24 h after OHT treatment. To control for OHT-related gene expression, we compared proneural-factor-induced programs to DsRed-OHT-treated samples (Figures 2D–2P and S2F–S2L).

Both factors induced drastic transcriptional changes compared to control (Figures 2D and 2K; Data S1). Consistent with previous data (Masserdotti et al., 2015), the programs differed profoundly, with only 17.6% of the upregulated genes common between the Ascl1ERT2- and Neurog2ERT2-induced cascades (log2FC > 1 and padj < 0.01, Figure 2E). This is 5 times higher than the proportion of genes commonly upregulated by these factors in GM astrocytes (Masserdotti et al., 2015). Gene Ontology (GO) analysis of commonly upregulated genes highlighted genes involved in nervous system development and regulation of neurogenesis (e.g., *Insm1*, *Sox11*, *Sox4*, *Dll1*; Figure 2F; Data S1) and in transcriptional activity (e.g., *Rcor2*, *Ncor2*, *Cbfa2t3*, and many transcription factors [TFs]; Figure 2G; Data S1). When focusing on genes uniquely induced by Ascl1ERT2 (107) or Neurog2ERT2 (40) (log2FC > 1, padj < 0.01 for 1 TF and log2FC ≤ 0 for the other; Figures 2E, 2J, S2H, and S2I), we found the former to specifically regulate genes related to muscle activity (e.g., *Tnn1, Tnnt2*; Figure 2H; Data S1), in line with previous observations on Ascl1 regulating alternative fates (Lee et al., 2020; Treutlein et al., 2016). Conversely, Neurog2ERT2 specifically regulated genes related to cytokine signaling, axon guidance/cell adhesion (e.g., *Robo1, Cxcr4*; Figure 2I), and regulation of GTPase activity (e.g., *Chn2*, *Itga6*, *Ntrk3*; Figure 2I; Data S1).

Both factors triggered the significant downregulation of many genes (Figure 2K), of which 25.8% were common (log2FC < −1, padj < 0.01; Figure 2L). These were connected to vasoconstriction (e.g., *Avpr1a*, *Ednra*, *Edn1*; Figure 2M; Data S1), suggesting the repression of astrocyte-specific functions (Alfaro-Cervello et al., 2012; Filosa et al., 2016), and involved the re-organization of the extracellular matrix (Figure 2N; Data S1). Typical genes expressed in astrocytes were also downregulated (e.g., *Aqp4*, *Slc1a3*, *Fgfr3*; Figure S2J) (Weng et al., 2019), as well as some expressed in glial progenitors (e.g., *Pax6*, *Fabp7*, *Vimentin*; Figure S2K) (Treutlein et al., 2016). Stringent criteria for gene selection revealed a subset of genes uniquely downregulated by Ascl1ERT2 (40 genes; Figure 2L), associated to phosphorylation and actin organization (e.g., *Tec*, *Nrp1*, *Actr3b*, Figure 2O; Data S1); conversely, Neurog2ERT2 specifically downregulated genes (43; Figure 2L) were associated with purine metabolism (e.g., *Aldoa*, *Bcl2l1*, *Eno1b*, *Uqcrc1*; Figure 2P; Data S1). Taken together, Ascl1ERT2 and Neurog2ERT2 not only induced different neurogenic programs at an early stage, but they also suppressed different aspects of the starter cell identity.

We next explored the expression changes of TFs and chromatin remodeling factors because important in fate specification and being a significant fraction of the regulated genes (13.9% in Ascl1ERT2 and 12.5% in Neurog2ERT2; log2FC > 1, padj < 0.01; Figure S2L; Data S1). Among the common upregulated TFs (log2FC > 1, padj < 0.01; Figure S2L), some have already been implicated in direct neuronal reprogramming (e.g., *Hes6*, *Insm1*, *Sox11*, *Sox4*, *Prox1*, *Trnp1*, *Zbtb18*, and *Ezh2*; Liu et al., 2018; Masserdotti et al., 2015; Wapinski et al., 2013). Interestingly, many other TFs were regulated specifically by Ascl1ERT2 (e.g., *Id3*, *Gata3*, *Klf10*, *Lhx3*; Figure 2Q) or Neurog2ERT2 (e.g., *Zbtb16*, *Sap18*, *Nhlh1*; Figure 2Q; Data S1), including *Gata3* and *Lhx3*, associated to V2 interneurons and motoneurons (Andrzejczuk et al., 2018; Misra et al., 2014). We also noted that both factors induced the expression of their respective endogenous genes (Data S1).

To identify important common TF- and chromatin-remodeler-specific downstream effectors, we constructed a gene expression regulatory network (Su et al., 2014a) by ranking TFs according to their induction, statistical significance and expression level (Figure 2R; see STAR Methods). The sub-network composed of shared genes (in yellow; Figure 2R) comprised many factors already found in cortical astrocyte reprogramming (e.g., Hes6, Sox11, Sox4, Atoh8, Olig1/2, Ncor2; Masserdotti et al., 2015), suggesting that this network may be common in different regions. Interestingly, the Ascl1ERT2-specific subnetwork was mainly composed of TFs (e.g., Sox8, Id1, Id2, Id3, Sox4, Sox11; Figure 2R, left), while the Neurog2ERT2-specific subnetwork was associated with many chromatin modifiers (e.g., Hmgb2, Smarcc1, Smarce1, Arid1a; Figure 2R, right) with Tcf4, the heterodimeric partner of Ascl1 and Neurog2 (Wang and

Figure 2. Transcriptional changes upon Ascl1ERT2- and Neurog2ERT2-induced direct reprogramming at early stages

(A) Scheme of the experimental design. Red bars indicate the time of OHT treatment.

(B) Micrographs of cells expressing Neurog2ERT2 and Ascl1ERT2 in the absence (upper panel) or presence (lower panel) of OHT. Scale bar: 100 μm.

(C) Barplot of %β3-tub+/DsRed+ cells with neuronal morphology following Neurog2ERT2 and Ascl1ERT2 activation with OHT at 8DPI. Mean and CI (95%) are shown (n = 3 for Ascl1ERT2 untreated; n = 4 for Neurog2ERT2 untreated; n = 6 for all the others). Linear regression model applied. ***p < 0.0001.

(D–P) Gene expression analysis showing bar plots (D and K) and Venn diagrams (E and L) of the number of genes differentially up- (D) or downregulated (K) (padj < 0.01 and log2FC > 0.5 or < −0.5) upon Ascl1ERT2 and Neurog2ERT2 induction after 24 h and the associated top GO terms as indicated (F–I for upregulated; M–P for downregulated) related to biological process (BP; F and M) or molecular function (MF; G and N) from commonly regulated genes (E and L) or specifically by Ascl1ERT2 (H and O) or Neurog2ERT2 (I and P). In (J), a heatmap is shown of most differentially regulated genes as indicated on the left side (padj < 0.01 and absolute log2FC > 1).

(Q) Scatterplot comparing Neurog2ERT2- and Ascl1ERT2-regulated genes (padj < 0.01 and log2FC > 2); transcription factors (TFs) in bold (blue, Ascl1 specific; black, Neurog2 specific).

(R) Network of most relevant TFs and chromatin modifiers induced after 24 h by Neurog2ERT2 (orange), Ascl1ERT2 (blue), and common (yellow). Bigger nodes correspond to the reprogramming factors; only connected nodes are plotted. Line width corresponds to StringDB scores.

(S) Example pathways identified after GSEA in Neurog2ERT2 and Ascl1ERT2 differentially expressed genes (Data S1).

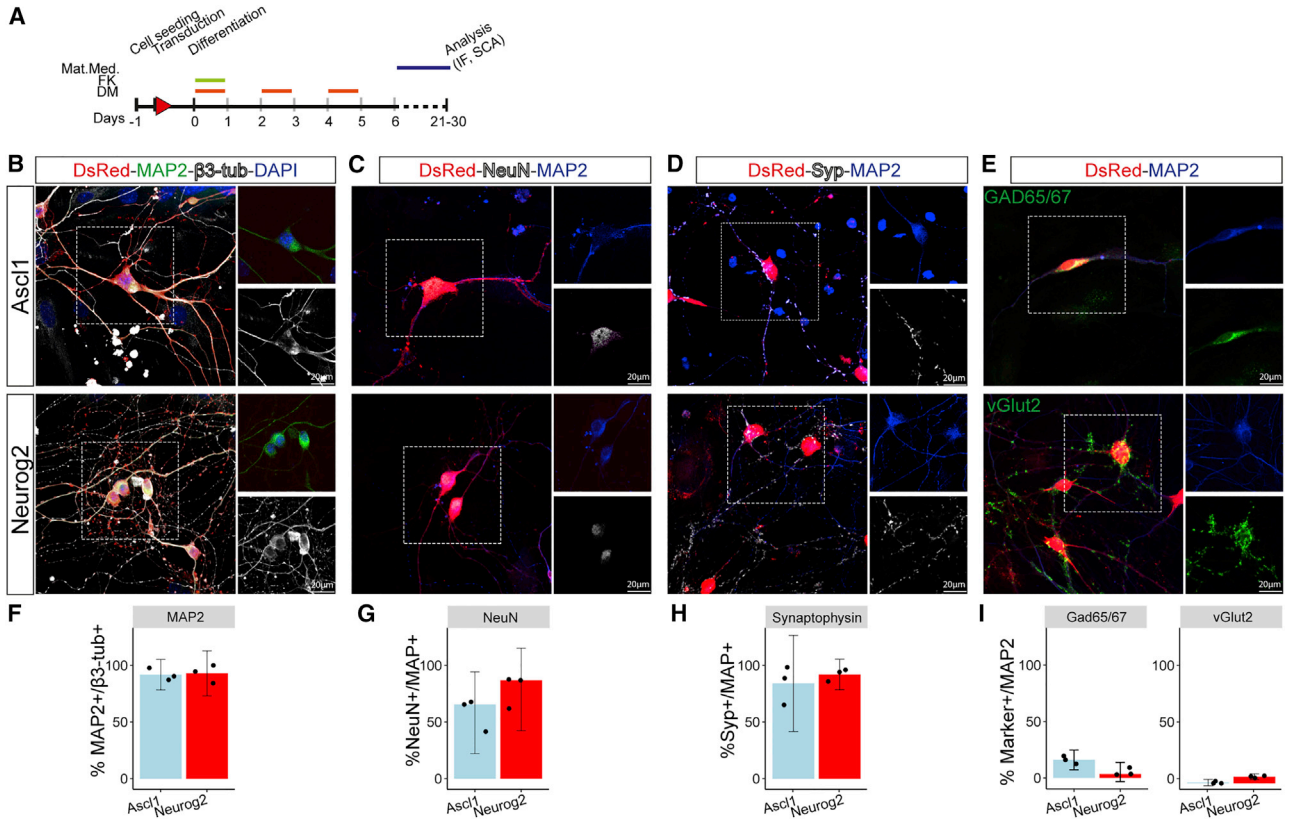


Figure 3. Expression of neuronal markers in Ascl1 and Neurog2 iNs

(A) Scheme of the experimental design. IF, immunofluorescence; SCA, single-cell analysis. Bars indicate the day of treatment with different small molecules.

(B–E) Micrographs depicting pan-neuronal, GABAergic, and glutamatergic markers as indicated in Ascl1- (top) and Neurog2- (bottom) iNs at 21–24 DPI. Scale bar: 20 μm.

(F–I) Bar charts of percentage of mature neuronal markers (F–H) and GAD65/67 or vGlut2 (I) in Ascl1- and Neurog2-iNs as indicated. n = 3 independent experiments.

Baker, 2015), linking them (Quevedo et al., 2019). Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) supported this observation, as genes related to "chromatin binding" and "chromatin assembly and disassembly" were more enriched in Neurog2ERT2 than in Ascl1ERT2 differentially expressed genes (Figure 2S).

Together, these data indicated that Ascl1ERT2 and Neurog2ERT2 activity quickly induced very different neurogenic cascades, characterized by pan-neuronal and TF-specific downstream cascades, which contribute to establish the new neuronal identity.

Molecular characterization of Ascl1- and Neurog2-iNs

To explore the neuronal identity elicited by these factors in SC astrocytes, we analyzed reprogrammed cells at later stages (21–30 DPI, Figure 3A): most iNs (β3-tubulin+DsRed+) were positive for the mature pan-neuronal markers Map2 (Ascl1, mean = 91.77%, CI = 10.54; Neurog2, mean = 92.93%, CI = 15.74; Figures 3B and 3F), NeuN (Ascl1, mean = 58.43%, CI = 25.70; Neurog2, mean = 79.03%, CI = 25.63; Figures 3C and 3G), and synaptophysin (Ascl1, mean = 84.10%, CI = 33.22; Neurog2, mean = 91.99%, CI = 10.2; Figures 3D and 3H). Some iNs were positive for the GABAergic marker Gad65/67 (Ascl1, mean = 15.93%, CI = 6.12; Neurog2, mean = 5.218%, CI = 4.60; Figures 3E and 3I, upper micrograph), while very few were positive for the glutamatergic presynaptic marker vGlut2, visible as punctate staining along the processes (Ascl1, mean = 0.65%, CI = 1.96; Neurog2, mean = 5.83%, CI = 1.9; Figures 3E and 3I, lower micrograph), similar to previous observations (Hu et al., 2019).

To better investigate iNs' identity, we collected single cells at 21–28 DPI (Figure S3A) and subjected them to RNA-seq (Picelli et al., 2014). Morphometric analysis (Figures S3B–S3E) did not reveal significant differences between Ascl1- and Neurog2-iNs. Single-cell RNA-seq analysis (scRNA-seq) was performed on 33 iNs (48 collected cells), identified by the expression of DsRed, small soma and the presence of two or more long and thin processes (11 Ascl1-iNs, 22 Neurog2-iNs), and six astrocytes (Ascl1-transduced with a flat morphology). Principal component analysis (PCA) clearly separated astrocytes and iNs (Figure 4A), with Ascl1-iNs clustering farther away from non-reprogrammed astrocytes than from Neurog2-iNs (Figure 4A). In line with the criteria for neuronal selection, hundreds of genes were differentially expressed between iNs and non-reprogrammed astrocytes (padj < 0.01; Figures 4B and 4E; Data S2), with many upregulated genes common to Ascl1- and Neurog2-iNs (Figure 4C; GO in

Figure 4D; Data S2). Thus, Ascl1- and Neurog2-iNs become more similar after the initial induction of very different neurogenic programs (for specific differences, see Figures S3F and S3G). GSEA (Figures S3H and S3I; Data S2) confirmed the enrichment for genes involved in synapse formation and neuronal activity for both TF-mediated reprogramming cascades. The few genes commonly downregulated (Figure 4F) were involved in detoxification (Figure 4G; Data S2) and β-oxidation (GSEA; Figures S3J and S3K; Data S2), well-known astrocyte hallmarks (Russo et al., 2020).

Interestingly, Neurog2-iNs were widely dispersed in the PCA plot, with some close to non-reprogrammed astrocytes (Figure 4A). As also the number of significantly downregulated genes was lower in Neurog2- than in Ascl1-iNs (Figure 4E), we examined whether this was due to an incomplete conversion. We generated an ''astrocyte score'' and a ''neuronal score'' (Astro-Score and Neuron-Score, respectively) by summing the normalized expression values of 68 markers for astrocytes (merging those used in Tripathy et al. [2018] and those identified in Weng et al. [2019]; averaged expression in Figure S3L) and 21 neuronal markers (including neuronal vesicles coding genes; averaged expression in Figure S3M). Ascl1-astrocytes (black dots, Figure 4H) were distinct from Ascl1-iNs; however, some Neurog2-iNs were close to the Ascl1-astrocytes, thus reflecting a rather incomplete conversion. Of note, Ascl1-iNs also retained the expression of some markers for astrocytes (e.g., *Fgfr3*, *Nfia*; Figure S3L), thus indicating the persistence of some aspects of the starter cell identity.

To determine the potential cause of the unsuccessful conversion, we examined the expression of endogenous *Ascl1* and *Neurog2*, as they were upregulated at an early stage by the forced expression of Ascl1 and Neurog2 (Data S1). Endogenous *Neurog2* was expressed in all Neurog2-iNs at a similar level (Figure 4I) and, likewise, *Ascl1* in Ascl1-iNs. However, a significant proportion of Neurog2-iNs also expressed high levels of *Ascl1* (Figure 4I), and these cells were closer to non-reprogrammed astrocytes (Figure S3N); conversely, *Neurog2* was barely detected in Ascl1-iNs. This correlated with the high expression of the markers for astrocytes *Slc1a3* and *Sox9* (Figure 4I) but had no effect on the neuronal program, as *Snap25* and *Syp* were similarly expressed. Thus, the induction of endogenous *Ascl1* in Neurog2-iNs might perturb the neuronal conversion and maintain some glial traits, as Ascl1 is also involved in the generation of glial lineages (Kelenis et al., 2018).

To determine the neuronal subtype identity elicited by these TFs, we analyzed the expression of genes characteristic for glutamatergic or GABAergic neurons and their respective receptors (Figures 4L and 4M). This revealed a surprisingly similar transmitter identity between the iNs induced by Ascl1 and Neurog2: both expressed some glutamatergic (Figure 4J) and GABAergic markers (Figure 4K) and their receptors (Figures 4L and 4M) (Häring et al., 2018). However, they differed in the expression of specific neuropeptides, such as galanin (*Gal*) in the Ascl1-iNs (Figure S3O), or in specific receptors (e.g., *Cnr1*, *Hrh3*, *Adra2a*, and *Chrna3* in Neurog2-iNs; *Htr5b* in Ascl1-iN; Figures S3P–S3R) (Ren et al., 2019).

Taken together, no clear-cut difference in the neuronal identity was detectable in Ascl1- versus Neurog2-iNs at later stages. However, this may be due to the heterogeneity in reprogramming, a surprising finding given the rather homogeneous maturity observed by electrophysiology (Figure 1). We therefore combined these modes of analysis using patch sequencing (patch-seq) to investigate the degree of correlation between electrophysiological and transcriptional maturity of iNs and their neuronal subtype identity.

Molecular analysis of electrophysiologically characterized Ascl1- and Neurog2-iNs using patch-seq

To directly correlate gene expression and electrophysiology, transduced SC-derived astrocytes were subjected to patch-seq (Cadwell et al., 2017; Winterer et al., 2019) from 14 DPI onward. First, we confirmed the high similarity of Ascl1- and Neurog2-iNs over DsRed-control (as in Figure 4C)—irrespective of their electrophysiological properties (Figure S4E; Data S3)—identifying only 72 genes specific for Ascl1 (Figures S4E and S4F; GO terms in Figure S4J; Data S3) and 51 upregulated only in Neurog2-iNs (Figures S4E and S4G; GO terms in Figure S4K; Data S3). Markers of astrocytes (e.g., *Aqp4*, *Nfia*, *Sox9*) were still expressed, though at lower levels than in DsRed-transduced astrocytes (Figure S4H).

Next, we grouped the cells based on their electrophysiological properties resulting in four groups (Figures S4A–S4D; STAR Methods): a control group (DsRed-transduced astrocytes; Figures 5A and 5B) and three classes of iNs. The different classes of iNs could be characterized by either high resting membrane potential and no spike or off-spike occurring after the depolarization (class 1 iNs; Figures 5C and 5D), small overshooting and immature spike (class 2 iNs; Figures 5E and 5F), or overshooting and (repetitive) firing (class 3 iNs; Figures 5G and 5H). Control cells and iNs were separated in the PCA plot (Figure 5I), with no detectable clustering among the diverse iN subgroups, as even class 1 and class 3 were intermingled (Figure 5I). This suggests that the electrophysiological properties of neurons may be determined mostly at the post-transcriptional level and/or by a small part of the transcriptional program. Differential expression analysis compared to control (log2FC > 1; padj < 0.01; Data S4) revealed 747 differentially expressed genes in class 1 iNs, 1,351 in class 2 iNs, and 1,384 in class 3 iNs, with 712 regulated in all and 1,298 co-regulated in the class 2 and class 3 iNs (Figure 5J). Genes differentially regulated in both class 2 and class 3 iNs were associated with synaptic activity (Data S4 for GO terms). Noting the presence of potassium channel (K-channel) subunit coding genes (e.g., *Kcnb1*, *Kcns2*; GO in Figure 5N; Data S4) among the molecular signature of class 3 iNs (Figure 5K) and the expression of glutamate receptor subunit coding genes (e.g., *Grik3*, *Grm2*; Figure 5L; GO Figure 5M; Data S4) in class 2 iNs signature (Figure 5L), we analyzed their expression in more detail (Figures 5O and S4L). Overall, both K-channels and glutamate receptors were more expressed in class 2 and class 3 iNs than in class 1 iNs and control astrocytes. Likewise, glutamatergic and GABAergic markers were expressed more in class 2 and class 3 iNs (Figures S4M and S4N). Some K-channels were expressed in both class 2 and class 3 iNs (e.g., *Kcnc3*), while others seemed specific to one of the proneural factors (e.g., *Kcnf1* for Neurog2; *Kcns2* for Ascl1; Figure 5O); glycinergic markers were barely detected (Figure S4O) (Callahan et al., 2019). Taken together,

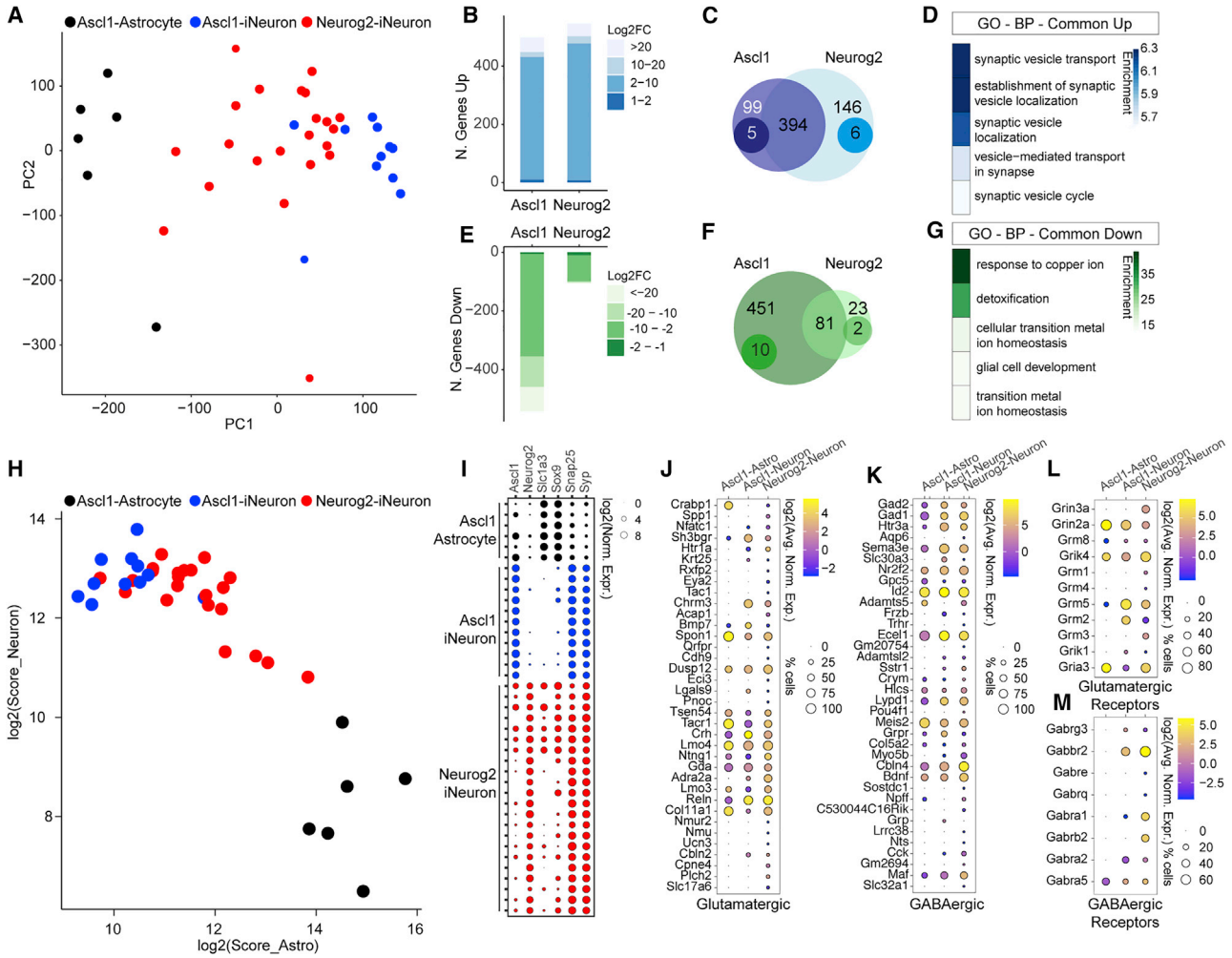


Figure 4. scRNA-seq of Ascl1- and Neurog2-iNs

(A) PCA of the 39 single cells isolated by patching. Neurog2-iNs (red dots) are localized between Ascl1-astrocytes (black dots) and Ascl1-iNs (blue dots).

(B, C, E, and F) Barplots (B and E) and Venn diagrams (C and F) of the number of genes upregulated (B and C) or downregulated (E and F) (padj < 0.01) between iNs and astrocytes. In (C) and (F), genes were considered if log2FC > |1| and padj < 0.01. Unique genes (darker circles) were considered if, for example, log2(FC-Ascl1) > 1, padj < 0.01, and log2(FC-Neurog2) ≤ 0, or, for example, log2(FC-Neurog2) > = 1, padj < 0.01, and log2(FC-Ascl1) ≤ 0.

(D and G) Top 5 GO terms (BP) from commonly upregulated (D) or downregulated (G) genes. Colors depict the enrichment over expected genes.

(H) Scatterplot depicting each cell based on log2(Astrocyte score) and log2(Neuronal score).

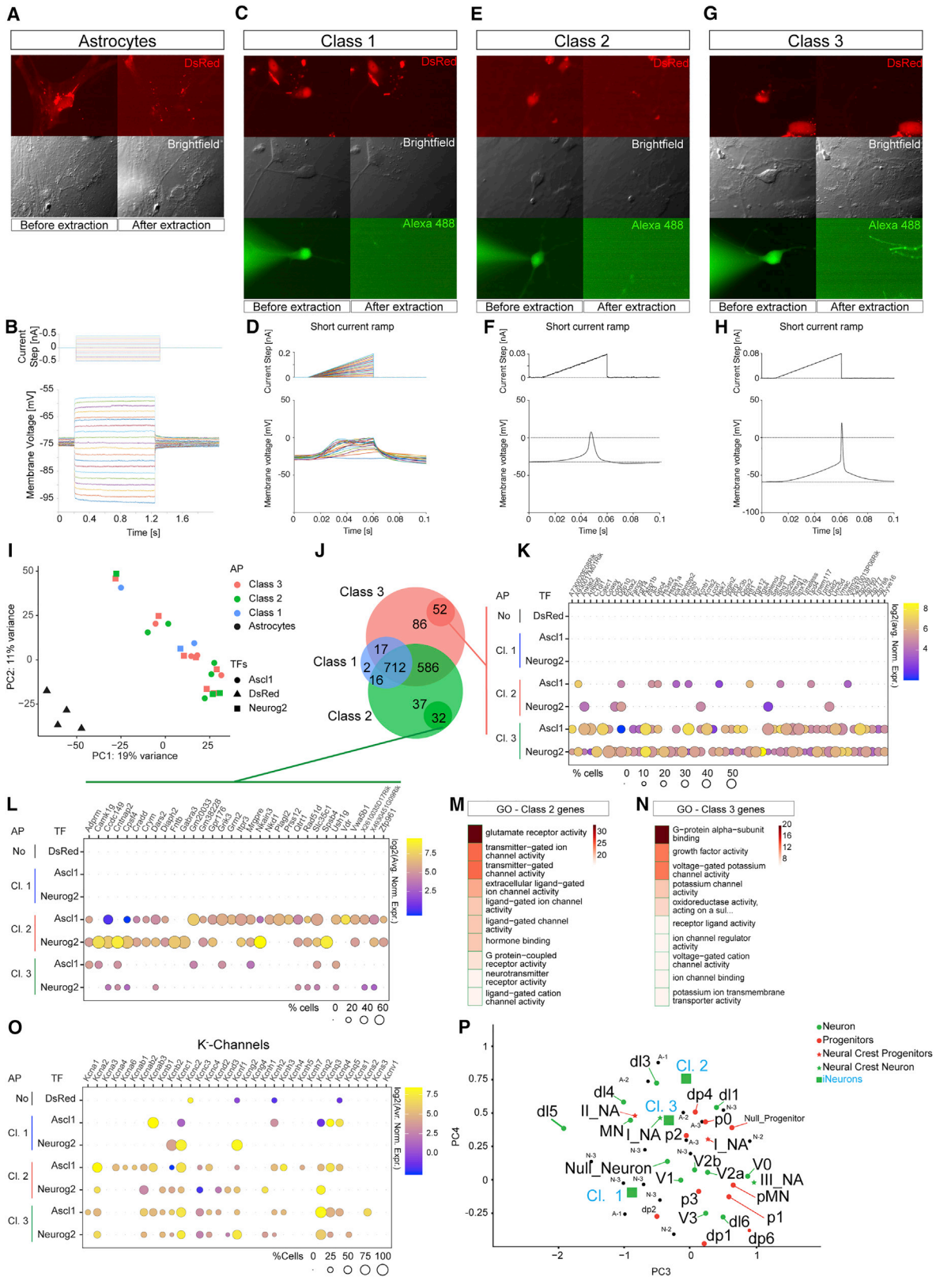
(I–M) Bubble plots showing the log2(expression) of selected genes (I), glutamatergic (J), GABAergic (K), and the respective receptors (L and M) in single cells. Size of the circles reflects the proportion of cells expressing the marker in each category (in J–M); color depicts the log2(average of normalized expression) (in J–M).

electrophysiology clearly identifies functional neuronal properties but cannot account for the transcriptional fidelity of conversion. Transcriptome analysis reveals the molecular state of the converted cells but cannot attribute functional features to neurons. Notably, the expression of some K-channels correlates with some functional properties of iNs.

Next, we compared our dataset with a published dataset of SC neurons (Delile et al., 2019). Control astrocytes were separated from neurons (Figure S4P); a closer look revealed that only the class 3 iNs appeared in the center of the cloud of endogenous SC neurons, while class 1 and class 2 iNs were at the margins (Figure 5P). Class 3 iNs were closer to V2b interneurons and p2, their progenitors (Karunaratne et al., 2002; Li et al., 2005). Consistent with this, Neurog2- and Ascl1-firing iNs were very close together, with some Ascl1 class 3 iNs closer to the p2 progenitors and some Neurog2 class 3 iNs closer to the V2b interneurons. Thus, both TFs elicited a rather similar neuronal subtype identity in SC-derived astrocytes, in pronounced contrast to previous observations following the conversion of astrocytes from other regions (Heinrich et al., 2010; Hu et al., 2019).

Region-specific identity markers are maintained *in vitro*

The above results clearly showed the generation of different iNs in the SC compared to cortex and midbrain astrocyte reprogramming (Heinrich et al., 2010; Liu et al., 2015). One possible cause for this may be the regionalized gene expression in the



(legend on next page)

starter cells. To determine this, we analyzed the transcriptome of MACS astrocytes from the cortical GM and the SC, immediately after MACS (acute) or cultured for 7 days under identical conditions (culture; Figure 6A). PCA separated the samples according to the region (PC1) and acute isolation versus culture (PC2; Figure S5A). Comparison between SC and GM astrocytes revealed similar differences in gene expression for both acutely isolated and cultured astrocytes (Figures 6B and 6C; Data S5). Known patterning genes from development were among the most differentially expressed genes in each condition analyzed (e.g., *Otx1*, *Emx2*, *Foxg1* in GM astrocytes; *Hox* genes in SC astrocytes; Figure S5B) (Hébert and Fishell, 2008; Philippidou and Dasen, 2013; Sagner and Briscoe, 2019). Indeed, GO term analysis supported the forebrain origin of GM astrocytes (Figures 6D and 6F; Data S5) and the SC signature for SC astrocytes (e.g., Hox genes; Figures 6E, 6F, 6H, and 6I; Data S5) from both acutely isolated and in vitro cells.

Interestingly, the transcriptional profile of cultured GM astrocytes was enriched for vasculature-related genes (Figures 6G and 6I; Data S5), supporting the notion of astrocytes as important players for blood-brain barrier formation (Langen et al., 2019). Of note, primary cultures of astrocytes from both regions showed high similarity with their acutely isolated counterparts (Figures S5C and S5F), with differences related to angiogenesis (*in vivo* for both GM and SC; Figures S5D and S5G; Data S6) and wound healing (Figures S5E and S5H; Data S6). Furthermore, acutely isolated cells, either from cortical GM or SC, expressed higher levels of astrocyte markers than did their cultured counterparts (Figures S5I and S5J). Importantly, the expression of region-specific patterning genes was maintained in iNs (Hox genes; Figure 6J), thus supporting the concept that neuronal identity is influenced by the regionalization of the starter astrocyte population.

To investigate the programs in astrocytes cultured from GM and the SC, we identified differentially expressed TFs, chromatin modifiers, and RNA-binding proteins and constructed a gene expression regulatory network (Su et al., 2014a). Three distinct branches emerged in SC astrocytes (Figure 6K): one related to RNA-binding proteins (red circles) (Matera and Wang, 2014), another comprising genes for neural tube development (e.g., *Pax6*, *FoxP1*, green circles), and the third including chromatin-related factors (e.g., *Ssrp1*, *Hmgb2*, *Dnmt1*, blue circles). Conversely, network analysis on genes from GM astrocytes revealed one ramified network (Figure 6L), in which *FoxG1*, *Mef2c*, and *Cebpb* seemed to be main hubs. GSEA highlighted that SC astrocytes express "dorso-ventral neural tube patterning" genes (Figures 6M and 6N; Data S5), and GM astrocytes express a higher number of "forebrain regionalization" genes (Figures 6O and 6P). In summary, both acutely isolated and *in vitro* cultured GM and SC astrocytes exhibited transcriptome differences and retained specific developmental and patterning hallmarks and gene networks (Figures 6 and S5), with high relevance for specifying neuronal subtypes.

Common and region-specific cascades induced by Ascl1 and Neurog2

As patterning genes are retained in iNs (Figure 6J), possibly influencing the reprograming, we compared Ascl1ERT2- and Neurog2ERT2-induced neurogenic cascades in SC- (Figure 2) and GM-derived astrocytes, obtained from published microarray data (Masserdotti et al., 2015) and induced for the same time (24 h) without any small-molecule treatment. This revealed a profound influence of the starter cells on the induced programs: most regulated genes were specific for a specific combination of starter cell and TF (Figure 6Q), with only 27.3% of Ascl1ERT2- and 12% of Neurog2ERT2-upregulated genes common between the astrocytes isolated from different regions (Figure 6Q). Irrespective of the starter cell, a large fraction of genes upregulated by Ascl1ERT2 (176 genes; Figures 6Q, yellow background, and S5K) were associated with neuronal functions (Figure 6R); among them, genes such as *Calm1*, *Pvalb*; the transcription factors *Lhx3*, *Lmo1*, and *Sox8*; and the signaling molecules *Bmp2*, *Bmp7*, *Sema6b*, and *Wnt9a* (Figure S5K). Neurog2ERT2 regulated 22 genes common to both regions (Figure 6Q, green background), among which were several synapse-associated protein coding genes, such as *Cnr1*, *Cplx2*, *Lrrtm1*, and *Cadm3* (Figure S5L). Overall, 20 genes (Figures 6Q and 6S, blue background) were commonly induced by both proneural factors from the different types of astrocytes, including many transcription factors (*Hes6*, *Insm1*, *Neurod4*, *Prox1*, *Sox11*, and *Trnp1*) (Masserdotti et al., 2015). Thus, the starter astrocyte regional identity profoundly influences the reprogramming process for each proneural factor with few pan-neuronal core regulators.

DISCUSSION

Direct conversion of SC-derived astrocytes showed that Ascl1 and Neurog2 elicit functionally mature neurons, with a signature

Figure 5. Patch-seq analysis of Ascl1- and Neurog2-iNs

(A, C, E, and G) Epifluorescent and brightfield pictures of cells selected for electrophysiological and transcriptome analysis. Shown are a control astrocyte (A; n = 4), class 1 non-firing reprogrammed cells (C; Ascl1 n = 2, Neurog2 n = 1), class 2 immature firing iNs (E, Ascl1, n = 5; Neurog2, n = 3), and class 3 firing iNs (G; Ascl1, n = 4; Neurog2, n = 7).

(B, D, F, and H) Examples of current steps (upper plots) and the corresponding membrane voltage responses (lower plots) for the cell types indicated in (A), (C), (E), and (G). For the astrocytes (B), long current step pulses are shown with hyperpolarizing and depolarizing steps. For the reprogrammed cells, short current ramp steps are shown with either all traces (class 1, D) or single traces (class 2, F; class 3, H).

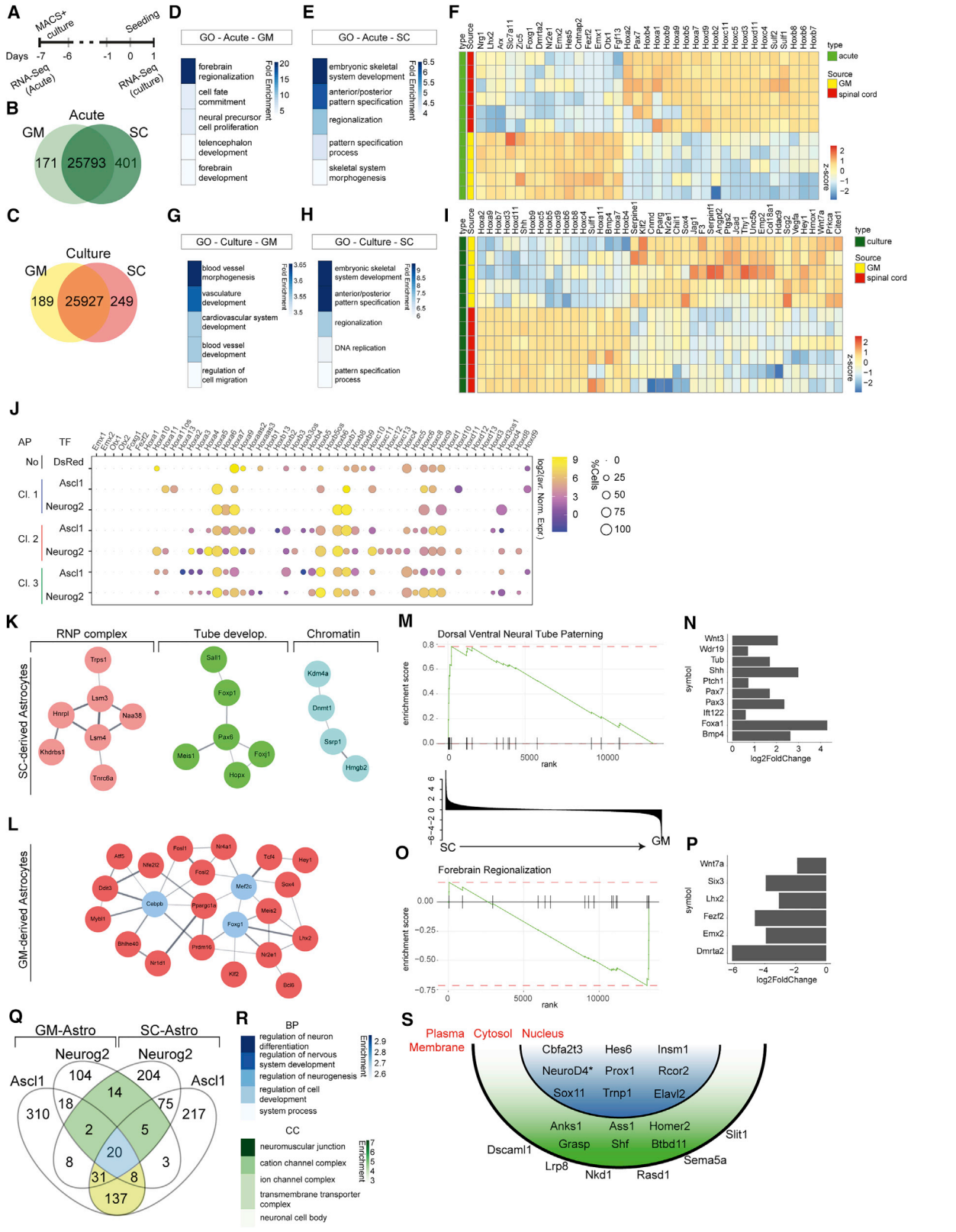
(I) PCA plot of 26 cells analyzed by patch-seq. Control cells (black triangles) are separated from iNs.

(J) Venn diagram of the genes shared among different classes of iNs.

(K, L, and O) Bubble plots showing the expression of genes highly expressed in class 3 (K) or class 2 (L) iNeurons and K-channels (O) in the different subgroups.

(M and N) Top 10 GO terms (MF) from highly expressed genes in class 2 (M) or class 3 iNs (N). Colors depict the enrichment over expected genes.

(P) PCA of iNs (clustered according to classes in C, E, and G) compared to SC progenitors and neurons from Delile et al. (2019). Each black dot represents an iN. A-n, Ascl1-Class n; N-n, Neurog2-Class n.



(legend on next page)

reminiscent of ventral SC interneuron identity, despite initially inducing very different transcriptional cascades. The distinct neuronal fate instructed from SC versus cortex astrocytes unraveled the potent influence of the regional identity of astrocytes, shown here to be maintained *in vitro*. These data further support the concept that astrocytes are well suited as starter cells for neuronal reprogramming: their regional specification and the programs elicited by the proneural factors may allow achieving the correct regional neuronal identity. Finally, our patch-seq data demonstrate the need for both electrophysiology and transcriptomics to ensure adequate conversion and fate acquisition.

Distinct molecular programs elicited by Ascl1 and Neurog2 in SC astrocytes

Using the inducible forms of Ascl1 and Neurog2, we identified the early induction of two largely distinct molecular programs (Figure 2), possibly also influenced by differences in the timing of the reprogramming or in the reprogramming efficiency. These programs comprise many TFs (Figure 2Q): of note, Neurog2ERT2 induced more chromatin remodeling factors than Ascl1ERT2 (Figure 2R), suggesting that Neurog2 may require specific co-factors and complexes to promote and stabilize the neuronal cascade (Figures 2R and 2S). Conversely, Ascl1 can apparently act as on-target pioneer factor (Wapinski et al., 2013) either alone or by recruiting endogenously expressed co-factors. Both TFs repressed astrocyte-specific genes (Figure S2I), while no neural stem cell marker was induced (Figure S2J), supporting the direct conversion of astrocytes without passing through a more stem-cell-like state, as is the case in fibroblast-to-neuron direct conversion (Treutlein et al., 2016), or upon the forced expression of *Sox2* in pericytes (Karow et al., 2018).

It was surprising to see no clear specification toward a GABAergic or glutamatergic identity (Figure 3) of iNs, contrary to recent reports (Hu et al., 2019) and different from cortical GM astrocytes reprogramming (Heinrich et al., 2010). Both Ascl1 and Neurog2 iNs from SC astrocytes were close to a ventral SC GABAergic V2b interneuron type (Figure 5P), consistent with the expression of *Gad1* and *Gad2* in some firing Ascl1 or Neurog2 iNs (Figure S4N). While most iNs did not fully complete the GABAergic neurotransmitter identity (Figure 3I), our results reveal that Ascl1 and Neurog2 do not instruct GABAergic versus glutamatergic neurons, as in the cortex astrocytes (Heinrich et al., 2010) and, to some extent, the midbrain astrocytes (Liu et al., 2015). Rather, the crosstalk between the reprogramming factors and the cellular context shapes the neuronal outcome: the persistence of patterning TFs in astrocytes of the respective regions may influence the developmental role of the proneural factors. During SC development, V2a and V2b interneurons derive from p2 progenitors, which express Ascl1, Neurog1, and Neurog2 in a mosaic pattern. Ascl1 promotes a V2a progenitor fate and Neurog2 inhibits a V2a progenitor fate, thereby promoting V2b neuronal identity (Misra et al., 2014). The acquired fate elicited by reprogramming closely reflects the role of these factors in SC development and may explain why some Neurog2-iNs can progress a bit further toward the V2b neuronal identity. Thus, Neurog2 can also induce GABAergic neuronal subtypes depending on its developmental role and the regional identity of the starter cells.

Our analysis further showed that Ascl1 and Neurog2 can interfere with each other, as shown for V2b interneuron fate acquisition (Misra et al., 2014) and in the telencephalon (Fode et al., 2000; Kovach et al., 2013; Schuurmans et al., 2004). Furthermore, Ascl1 is also involved in gliogenesis throughout the CNS, including the SC (Kelenis et al., 2018; Vue et al., 2014). As such, the co-expression of endogenous *Ascl1* and *Neurog2* in some Neurog2 iNs correlated with an incomplete repression of the astrocytic signature (Figures 4I, S3L, and S4H). However, *Ascl1* and *Neurog2* are co-expressed at E12.5 in the dorsal telencephalic ventricular zone (Britz et al., 2006), and double-knockout (KO) embryos showed specific defects in neurogenesis (Dennis et al., 2017; Nieto et al., 2001), further supporting their context-dependent activity; as such, in fibroblasts their co-expression rather improves reprogramming into neurons (Herdy et al., 2019; Ladewig et al., 2012). Thus, the function of these TFs is context dependent despite their common neurogenic role: astrocytes provide a context that allows predictions of the induced fate based on CNS development from radial glial cells.

The contribution of regionalization to the reprogramming process

Developmentally instructed patterning information was present in acutely isolated astrocytes and maintained *in vitro*, with GM-derived astrocytes expressing TFs characteristic of the dorsal telencephalon (e.g., *Emx1*, *Otx1*) and SC-derived astrocytes expressing *Hox* genes (Figure S5B). Thus, the maintenance of regionalization contributes to the induction of different neurogenic

Figure 6. Transcriptional differences between astrocytes from SC and cerebral cortex GM

(A) Scheme of sample collection for RNA-seq.

(B and C) Venn diagram of the transcriptome of GM and SC astrocytes acutely isolated (B) or following 7 days in culture (C).

(D–H) Top 5 GO terms (BP) associated with genes differentially expressed in acutely isolated GM (D) and SC (E) and cultured GM (G) and SC (H) astrocytes.

(F and I) Heatmaps of the relative expression of genes comprised in the top GO for each region in acute (F) and cultured condition (I).

(J) Bubble plot depicting the percentage of cells and log2(averaged normalized expression) of patterning genes in different classes of iNs (dataset from Figure 5).

(K and L) Network analysis of TFs, enriched chromatin modifiers, and RNA-binding protein in cultured SC astrocytes (K) or cultured GM astrocytes (L) with main hubs in blue.

(M and O) Examples of pathways identified by GSEA after comparing cultured GM and SC astrocytes transcriptome. SC neural tube patterning genes (M); GM-related forebrain regionalization genes (O).

(N and P) Barplots depicting the log2FC of genes associated with (M) (shown in N) or with (O) (shown in P).

(Q) Overlap between the genes induced by Ascl1ERT2 and Neurog2ERT2 in GM (log2FC > 1.3 and $p < 0.05$) and SC (log2FC > 2 and padj < 0.01) astrocytes at 24 h.

(R and S) Top 5 GO terms (R; cellular compartment, CC) and sketch of the cellular localization (S) of 20 commonly upregulated genes. *Neurod4* is regulated in SC-derived astrocytes by Ascl1ERT2 only with padj = 0.06.

cascades by the same TFs in astrocytes from different regions (Figure 6Q), resulting in the generation of different neurons, as reported for different brain regions including distinct thalamic nuclei *in vitro* and *in vivo* (Herrero-Navarro et al., 2021; Mattugini et al., 2019; Qian et al., 2020; Zhou et al., 2020). Of note, GM astrocytes could be more efficiently reprogrammed than SC astrocytes, suggesting a broader effect of regionalization on direct conversion. For example, the transcriptional and proteomic context in different astrocytes could regulate proneural gene activity (e.g., via the expression of specific cofactors or via their phosphorylation; Ali et al., 2014; Hindley et al., 2012), a very relevant aspect for *in vivo* repair. Only few genes were induced by both TFs in both regions (e.g., *NeuroD4*, *Hes6, Insm1*, *Prox1*, and *Sox11*; Figures 6Q and 6R) (Masserdotti et al., 2015), thus possibly representing the "pan-neurogenic" core network sufficient to instruct a neuronal fate in different cell types, including astrocytes from other CNS regions.

In summary, these data provide compelling evidence for a major contribution of the starter cell in shaping the ability of Ascl1 and Neurog2 to exploit their reprogramming potential, which calls for the identification of starter-cell-specific cocktails of reprogramming factors. Furthermore, it suggests that many reprogrammed cells retain features of their original identity, prompting the need for molecular characterization of the final neuronal outcome.

Patch-seq reveals low correlation between electrophysiology and transcriptome of single iNs

The gold standard of iNs is electrophysiology, and rightly so, as firing and synaptic connectivity are central to their functional roles. However, thus far, the electrophysiological and transcriptional state of single iNs using patch-seq was not assessed. Here, we show that there is no correlation in the overall clustering of different electrophysiological classes of iNs according to their transcriptome, suggesting that these major functional differences might depend on only few genes (which, hence, would not influence the clustering) or on other aspects (e.g., post-translational modifications). Indeed, channels and receptors have a long half-life as proteins and often do not appear as distinguishing features in scRNA-seq data. Therefore, the transcriptome cannot predict the electrophysiological state of iNs, even though we unraveled the higher expression of some K-channels in the most mature iNs (Figures 5N and 5O). These contribute to maintain the resting membrane potential and repolarize neurons, a distinctive feature of firing iNs. Importantly, electrophysiology cannot be used as a predictor of the overall fate conversion at the transcriptional level. Hence, it is essential to complement electrophysiology with transcriptional analysis as quality control for the identity of iNs, to avoid incomplete phenotypes especially for *in vivo* neuronal replacement.

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
  - Wild-type mice (primary cell culture, IHC, RNA-sequencing)
  - Primary cultures of astrocytes
- METHODS DETAILS
  - Plasmids and viral production
  - Transduction
  - Long term culture of reprogrammed astrocytes
  - Fluorescence-activated cell sorting
  - Immunocytochemistry
  - Bulk RNA sequencing
  - RNA-seq analysis
  - Single cell RNA-sequencing patch-seq related to Figures 4 and S3
  - Single cell-RNA-seq Patch-seq related to Figures 5 and S4
  - Single-cell morphometry
  - Ranking of transcription factors
  - Comparison of scRNA-seq with publicly available data
  - Electrophysiology
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109409.

ACKNOWLEDGMENTS

We are very grateful to the excellent technical assistance of Ines Mühlhahn, Detlef Franzen, Manja Thorwirth, and Andrea Steiner-Mezzadri (BMC, LMU); Judith Fischer for FACS training and troubleshooting; Tobias Straub, head of the Bioinformatic Unit at the BMC, for aligning RNA-seq raw data and advice on analysis; Jovica Ninkovic and Mike Myoga for critical comments; and LaFuga (Gene Center, Munich) and Next Generation Sequencing Core Facility (Helmholtz) for sequencing. This work was funded by the German Research Foundation grants SFB 870, SPP1757 (to M.G.), the advanced ERC ChroNeuroRepair (to M.G.), ERA-Net neuron grant MICRONET (to M.G.), the EU consortium NSC Reconstruct (M.G.), German Research Foundation grant TRR274 (to M.G.), and grant EN 1093/2-1 (to A.J. and W.E.). This work is dedicated to Rinaldo Masserdotti.

AUTHOR CONTRIBUTIONS

G.M. conceived and designed the project. J.K. and K.K. performed the isolation, culturing, and most of the reprogramming experiments. J.K. and K.K. evaluated the reprogramming efficiency at 8DPI; J.K. performed and quantified the experiments at day 21. K.K. prepared the cells for FACS analysis and extracted RNA for bulk RNA-seq analysis. B.A.H. prepared primary cultures of astrocytes from cortex GM, collected single cells, processed the samples with SmartSeq2, and prepared the libraries for sequencing; D.P. collected single iNs for patch-seq, processed them via Smart-Seq2 protocol, and performed morphometric analysis. V.B performed patch-seq experiments, analyzed the electrophysiological data, prepared the samples via Smart-Seq2, and performed initial analysis of RNA-seq data. A.J. prepared bulk-adapted mcSCRB-seq libraries; T.S.-E. contributed to establish MACS protocol for SC, trained J.K. and K.K., and performed immunocytochemistry; T.R. performed whole-cell patch-clamp experiments and analyzed the electrophysiological properties of iNs. W.E. provided reagents; G.M. performed FACS and analyzed data from bulk and single-cell RNA-seq; P.S. compared

single-cell RNA with publicly available data and analyzed single-cell RNA-seq with G.M.; M.G. contributed to the conceptual organization of the manuscript and the patch-seq data and financed the work. G.M. and M.G. wrote the manuscript, and all authors contributed corrections and comments.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 24, 2021
Revised: April 14, 2021
Accepted: June 24, 2021
Published: July 20, 2021; corrected online: August 6, 2021

REFERENCES

Abernathy, D.G., Kim, W.K., McCoy, M.J., Lake, A.M., Ouwenga, R., Lee, S.W., Xing, X., Li, D., Lee, H.J., Heuckeroth, R.O., et al. (2017). MicroRNAs Induce a Permissive Chromatin Environment that Enables Neuronal Subtype-Specific Reprogramming of Adult Human Fibroblasts. Cell Stem Cell *21*, 332–348.e9.

Alfaro-Cervello, C., Soriano-Navarro, M., Mirzadeh, Z., Alvarez-Buylla, A., and Garcia-Verdugo, J.M. (2012). Biciliated ependymal cell proliferation contributes to spinal cord growth. J. Comp. Neurol. *520*, 3528–3552.

Alexa, A., and Rahnenfuhrer, G. (2018). topGO: Enrichment Analysis for Gene Ontology, version 2.34.0 (R Foundation).

Ali, F.R., Cheng, K., Kirwan, P., Metcalfe, S., Livesey, F.J., Barker, R.A., and Philpott, A. (2014). The phosphorylation status of Ascl1 is a key determinant of neuronal differentiation and maturation in vivo and in vitro. Development *141*, 2216–2224.

Andrzejczuk, L.A., Banerjee, S., England, S.J., Voufo, C., Kamara, K., and Lewis, K.E. (2018). Tal1, Gata2a, and Gata3 Have Distinct Functions in the Development of V2b and Cerebrospinal Fluid-Contacting KA Spinal Neurons. Front. Neurosci. *12*, 170.

Bagnoli, J.W., Ziegenhain, C., Janjic, A., Wange, L.E., Vieth, B., Parekh, S., Geuder, J., Hellmann, I., and Enard, W. (2018). Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat. Commun. *9*, 2937.

Barker, R.A., Götz, M., and Parmar, M. (2018). New approaches for brain repair-from rescue to reprogramming. Nature *557*, 329–334.

Bean, B.P. (2007). The action potential in mammalian central neurons. Nat. Rev. Neurosci. *8*, 451–465.

Berninger, B., Costa, M.R., Koch, U., Schroeder, T., Sutor, B., Grothe, B., and Götz, M. (2007). Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. J. Neurosci. *27*, 8654–8664.

Borromeo, M.D., Meredith, D.M., Castro, D.S., Chang, J.C., Tung, K.C., Guillemot, F., and Johnson, J.E. (2014). A transcription factor network specifying inhibitory versus excitatory neurons in the dorsal spinal cord. Development *141*, 2803–2812.

Britz, O., Mattar, P., Nguyen, L., Langevin, L.M., Zimmer, C., Alam, S., Guillemot, F., and Schuurmans, C. (2006). A role for proneural genes in the maturation of cortical progenitor cells. Cereb. Cortex *16*, i138–i151.

Cadwell, C.R., Scala, F., Li, S., Livrizzi, G., Shen, S., Sandberg, R., Jiang, X., and Tolias, A.S. (2017). Multimodal profiling of single-cell morphology, electrophysiology, and gene expression using Patch-seq. Nat. Protoc. *12*, 2531–2553.

Callahan, R.A., Roberts, R., Sengupta, M., Kimura, Y., Higashijima, S.I., and Bagnall, M.W. (2019). Spinal V2b neurons reveal a role for ipsilateral inhibition in speed control. eLife *8*, e47837.

Chouchane, M., Melo de Farias, A.R., Moura, D.M.S., Hilscher, M.M., Schroeder, T., Leão, R.N., and Costa, M.R. (2017). Lineage Reprogramming of Astroglial Cells from Different Origins into Distinct Neuronal Subtypes. Stem Cell Reports *9*, 162–176.

Church, V.A., Cates, K., Capano, L., Aryal, S., Kim, W.K., and Yoo, A.S. (2021). Generation of Human Neurons by microRNA-Mediated Direct Conversion of Dermal Fibroblasts. Methods Mol. Biol. *2239*, 77–100.

Crochet, S., Fuentealba, P., Timofeev, I., and Steriade, M. (2004). Selective amplification of neocortical neuronal output by fast prepotentials in vivo. Cereb. Cortex *14*, 1110–1121.

Delile, J., Rayon, T., Melchionda, M., Edwards, A., Briscoe, J., and Sagner, A. (2019). Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. Development *146*, dev173807.

Dennis, D.J., Wilkinson, G., Li, S., Dixit, R., Adnani, L., Balakrishnan, A., Han, S., Kovach, C., Gruenig, N., Kurrasch, D.M., et al. (2017). *Neurog2* and *Ascl1* together regulate a postmitotic derepression circuit to govern laminar fate specification in the murine neocortex. Proc. Natl. Acad. Sci. USA *114*, E4934–E4943.

Filosa, J.A., Morrison, H.W., Iddings, J.A., Du, W., and Kim, K.J. (2016). Beyond neurovascular coupling, role of astrocytes in the regulation of vascular tone. Neuroscience *323*, 96–109.

Fode, C., Ma, Q., Casarosa, S., Ang, S.L., Anderson, D.J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. Genes Dev. *14*, 67–80.

Földy, C., Darmanis, S., Aoto, J., Malenka, R.C., Quake, S.R., and Südhof, T.C. (2016). Single-cell RNAseq reveals cell adhesion molecule profiles in electrophysiologically defined neurons. Proc. Natl. Acad. Sci. USA *113*, E5222–E5231.

Gascón, S., Murenu, E., Masserdotti, G., Ortega, F., Russo, G.L., Petrik, D., Deshpande, A., Heinrich, C., Karow, M., Robertson, S.P., et al. (2016). Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. Cell Stem Cell *18*, 396–409.

Golding, N.L., and Spruston, N. (1998). Dendritic sodium spikes are variable triggers of axonal action potentials in hippocampal CA1 pyramidal neurons. Neuron *21*, 1189–1200.

Guo, Z., Zhang, L., Wu, Z., Chen, Y., Wang, F., and Chen, G. (2014). In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. Cell Stem Cell *14*, 188–202.

Hartig, F. (2021). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models. R package version 0.4.3. http://florianhartig.github.io/DHARMa/.

Häring, M., Zeisel, A., Hochgerner, H., Rinwa, P., Jakobsson, J.E.T., Lönnerberg, P., La Manno, G., Sharma, N., Borgius, L., Kiehn, O., et al. (2018). Neuronal atlas of the dorsal horn defines its architecture and links sensory input to transcriptional cell types. Nat. Neurosci. *21*, 869–880.

Hébert, J.M., and Fishell, G. (2008). The genetics of early telencephalon patterning: some assembly required. Nat. Rev. Neurosci. *9*, 678–685.

Heinrich, C., Blum, R., Gascón, S., Masserdotti, G., Tripathi, P., Sánchez, R., Tiedt, S., Schroeder, T., Götz, M., and Berninger, B. (2010). Directing astroglia from the cerebral cortex into subtype specific functional neurons. PLoS Biol. *8*, e1000373.

Heinrich, C., Gascón, S., Masserdotti, G., Lepier, A., Sanchez, R., Simon-Ebert, T., Schroeder, T., Götz, M., and Berninger, B. (2011). Generation of subtype-specific neurons from postnatal astroglia of the mouse cerebral cortex. Nat. Protoc. *6*, 214–228.

Heins, N., Malatesta, P., Cecconi, F., Nakafuku, M., Tucker, K.L., Hack, M.A., Chapouton, P., Barde, Y.A., and Götz, M. (2002). Glial cells generate neurons: the role of the transcription factor Pax6. Nat. Neurosci. *5*, 308–315.

Henke, R.M., Savage, T.K., Meredith, D.M., Glasgow, S.M., Hori, K., Dumas, J., MacDonald, R.J., and Johnson, J.E. (2009). Neurog2 is a direct downstream target of the Ptf1a-Rbpj transcription complex in dorsal spinal cord. Development *136*, 2945–2954.

Herdy, J., Schafer, S., Kim, Y., Ansari, Z., Zangwill, D., Ku, M., Paquola, A., Lee, H., Mertens, J., and Gage, F.H. (2019). Chemical modulation of transcriptionally enriched signaling pathways to optimize the conversion of fibroblasts into neurons. eLife *8*, e41356.

Herrero-Navarro, Á., Puche-Aroca, L., Moreno-Juan, V., Sempere-Ferràndez, A., Espinosa, A., Susín, R., Torres-Masjoan, L., Leyva-Díaz, E., Karow, M., Figueres-Oñate, M., et al. (2021). Astrocytes and neurons share region-specific transcriptional signatures that confer regional identity to neuronal reprogramming. Sci. Adv. *7*, eabe8978.

Hindley, C., Ali, F., McDowell, G., Cheng, K., Jones, A., Guillemot, F., and Philpott, A. (2012). Post-translational modification of Ngn2 differentially affects transcription of distinct targets to regulate the balance between progenitor maintenance and differentiation. Development *139*, 1718–1723.

Hu, X., Qin, S., Huang, X., Yuan, Y., Tan, Z., Gu, Y., Cheng, X., Wang, D., Lian, X.F., He, C., and Su, Z. (2019). Region-Restrict Astrocytes Exhibit Heterogeneous Susceptibility to Neuronal Reprogramming. Stem Cell Reports *12*, 290–304.

Kantzer, C.G., Boutin, C., Herzig, I.D., Wittwer, C., Reiß, S., Tiveron, M.C., Drewes, J., Rockel, T.D., Ohlig, S., Ninkovic, J., et al. (2017). Anti-ACSA-2 defines a novel monoclonal antibody for prospective isolation of living neonatal and adult astrocytes. Glia *65*, 990–1004.

Karow, M., Camp, J.G., Falk, S., Gerber, T., Pataskar, A., Gac-Santel, M., Kageyama, J., Brazovskaja, A., Garding, A., Fan, W., et al. (2018). Direct pericyte-to-neuron reprogramming via unfolding of a neural stem cell-like program. Nat. Neurosci. *21*, 932–940.

Karunaratne, A., Hargrave, M., Poh, A., and Yamada, T. (2002). GATA proteins identify a novel ventral interneuron subclass in the developing chick spinal cord. Dev. Biol. *249*, 30–43.

Kelenis, D.P., Hart, E., Edwards-Fligner, M., Johnson, J.E., and Vue, T.Y. (2018). ASCL1 regulates proliferation of NG2-glia in the embryonic and adult spinal cord. Glia *66*, 1862–1880.

Kovach, C., Dixit, R., Li, S., Mattar, P., Wilkinson, G., Elsen, G.E., Kurrasch, D.M., Hevner, R.F., and Schuurmans, C. (2013). Neurog2 simultaneously activates and represses alternative gene expression programs in the developing neocortex. Cereb. Cortex *23*, 1884–1900.

Ladewig, J., Mertens, J., Kesavan, J., Doerr, J., Poppe, D., Glaue, F., Herms, S., Wernet, P., Kögler, G., Müller, F.J., et al. (2012). Small molecules enable highly efficient neuronal conversion of human fibroblasts. Nat. Methods *9*, 575–578.

Langen, U.H., Ayloo, S., and Gu, C. (2019). Development and Cell Biology of the Blood-Brain Barrier. Annu. Rev. Cell Dev. Biol. *35*, 591–613.

Lee, Q.Y., Mall, M., Chanda, S., Zhou, B., Sharma, K.S., Schaukowitch, K., Adrian-Segarra, J.M., Grieder, S.D., Kareta, M.S., Wapinski, O.L., et al. (2020). Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. Nat. Cell Biol. *22*, 401–411.

Li, S., Misra, K., Matise, M.P., and Xiang, M. (2005). Foxn4 acts synergistically with Mash1 to specify subtype identity of V2 interneurons in the spinal cord. Proc. Natl. Acad. Sci. USA *102*, 10688–10693.

Liu, M.L., Zang, T., Zou, Y., Chang, J.C., Gibson, J.R., Huber, K.M., and Zhang, C.L. (2013). Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. Nat. Commun. *4*, 2183.

Liu, Y., Miao, Q., Yuan, J., Han, S., Zhang, P., Li, S., Rao, Z., Zhao, W., Ye, Q., Geng, J., et al. (2015). Ascl1 Converts Dorsal Midbrain Astrocytes into Functional Neurons In Vivo. J. Neurosci. *35*, 9336–9355.

Liu, Y., Yu, C., Daley, T.P., Wang, F., Cao, W.S., Bhate, S., Lin, X., Still, C., 2nd, Liu, H., Zhao, D., et al. (2018). CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. Cell Stem Cell *23*, 758–771.e8.

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, D.C., Niu, T., and Alaynick, W.A. (2015). Molecular and cellular development of spinal cord locomotor circuitry. Front. Mol. Neurosci. *8*, 25.

Masserdotti, G., Gillotin, S., Sutor, B., Drechsel, D., Irmler, M., Jørgensen, H.F., Sass, S., Theis, F.J., Beckers, J., Berninger, B., et al. (2015). Transcriptional Mechanisms of Proneural Factors and REST in Regulating Neuronal Reprogramming of Astrocytes. Cell Stem Cell *17*, 74–88.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108–121.

Mattugini, N., Bocchi, R., Scheuss, V., Russo, G.L., Torper, O., Lao, C.L., and Götz, M. (2019). Inducing Different Neuronal Subtypes from Astrocytes in the Injured Mouse Cerebral Cortex. Neuron *103*, 1086–1095.e5.

Meyer, K., Ferraiuolo, L., Miranda, C.J., Likhite, S., McElroy, S., Renusch, S., Ditsworth, D., Lagier-Tourenne, C., Smith, R.A., Ravits, J., et al. (2014). Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. Proc. Natl. Acad. Sci. USA *111*, 829–832.

Misra, K., Luo, H., Li, S., Matise, M., and Xiang, M. (2014). Asymmetric activation of Dll4-Notch signaling by Foxn4 and proneural factors activates BMP/TGFβ signaling to specify V2b interneurons in the spinal cord. Development *141*, 187–198.

Mizuguchi, R., Kriks, S., Cordes, R., Gossler, A., Ma, Q., and Goulding, M. (2006). Ascl1 and Gsh1/2 control inhibitory and excitatory cell fate in spinal sensory interneurons. Nat. Neurosci. *9*, 770–778.

Nieto, M., Schuurmans, C., Britz, O., and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. Neuron *29*, 401–413.

Ory, D.S., Neugeboren, B.A., and Mulligan, R.C. (1996). A stable human-derived packaging cell line for production of high titer retrovirus/vesicular stomatitis virus G pseudotypes. Proc. Natl. Acad. Sci. USA *93*, 11400–11406.

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2018). zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience *7*, giy059.

Parras, C.M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D.J., and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. Genes Dev. *16*, 324–338.

Philippidou, P., and Dasen, J.S. (2013). Hox genes: choreographers in neural development, architects of circuit organization. Neuron *80*, 12–34.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181.

Qian, H., Kang, X., Hu, J., Zhang, D., Liang, Z., Meng, F., Zhang, X., Xue, Y., Maimon, R., Dowdy, S.F., et al. (2020). Reversing a model of Parkinson's disease with in situ converted nigral neurons. Nature *582*, 550–556.

Quevedo, M., Meert, L., Dekker, M.R., Dekkers, D.H.W., Brandsma, J.H., van den Berg, D.L.C., Ozgür, Z., van IJcken, W.F.J., Demmers, J., Fornerod, M., and Poot, R.A. (2019). Mediator complex interaction partners organize the transcriptional network that defines neural stem cells. Nat. Commun. *10*, 2669.

Rackham, O.J., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., Suzuki, H., Nefzger, C.M., Daub, C.O., Shin, J.W., et al.; FANTOM Consortium (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335.

Rao, Z., Wang, R., Li, S., Shi, Y., Mo, L., Han, S., Yuan, J., Jing, N., and Cheng, L. (2021). Molecular Mechanisms Underlying Ascl1-Mediated Astrocyte-to-Neuron Conversion. Stem Cell Reports *16*, 534–547.

Ren, J., Isakova, A., Friedmann, D., Zeng, J., Grutzner, S.M., Pun, A., Zhao, G.Q., Kolluru, S.S., Wang, R., Lin, R., et al. (2019). Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. eLife *8*, e49424.

Rivetti di Val Cervo, P., Romanov, R.A., Spigolon, G., Masini, D., Martín-Montañez, E., Toledo, E.M., La Manno, G., Feyder, M., Pifl, C., Ng, Y.H., et al. (2017). Induction of functional dopamine neurons from human astrocytes in vitro and mouse astrocytes in a Parkinson's disease model. Nat. Biotechnol. *35*, 444–452.

Russo, G.L., Sonsalla, G., Natarajan, P., Breunig, C.T., Bulli, G., Merl-Pham, J., Schmitt, S., Giehrl-Schwab, J., Giesert, F., Jastroch, M., et al. (2020). CRISPR-Mediated Induction of Neuron-Enriched Mitochondrial Proteins Boosts Direct Glia-to-Neuron Conversion. Cell Stem Cell *28*, 524–534.e7.

Sagner, A., and Briscoe, J. (2019). Establishing neuronal diversity in the spinal cord: a time and a place. Development *146*, dev182154.

Schuurmans, C., Armant, O., Nieto, M., Stenman, J.M., Britz, O., Klenin, N., Brown, C., Langevin, L.M., Seibt, J., Tang, H., et al. (2004). Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. EMBO J. *23*, 2892–2902.

Smith, D.K., Yang, J., Liu, M.L., and Zhang, C.L. (2016). Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. Stem Cell Reports *7*, 955–969.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell Stem Cell *9*, 205–218.

Su, G., Morris, J.H., Demchak, B., and Bader, G.D. (2014a). Biological network exploration with Cytoscape 3. Curr. Protoc. Bioinformatics *47*, 8.13.1-24.

Su, Z., Niu, W., Liu, M.L., Zou, Y., and Zhang, C.L. (2014b). In vivo conversion of astrocytes to neurons in the injured adult spinal cord. Nat. Commun. *5*, 3338.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Sergushichev, A.A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv. https://doi.org/10.1101/060012.

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. *45*, D362–D368.

Tai, W., Wu, W., Wang, L.L., Ni, H., Chen, C., Yang, J., Zang, T., Zou, Y., Xu, X.M., and Zhang, C.L. (2021). In vivo reprogramming of NG2 glia enables adult neurogenesis and functional recovery following spinal cord injury. Cell Stem Cell *28*, 923–937.e4.

Torper, O., Pfisterer, U., Wolf, D.A., Pereira, M., Lau, S., Jakobsson, J., Björklund, A., Grealish, S., and Parmar, M. (2013). Generation of induced neurons via direct conversion in vivo. Proc. Natl. Acad. Sci. USA *110*, 7038–7043.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., and Quake, S.R. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Tripathy, S.J., Toker, L., Bomkamp, C., Mancarci, B.O., Belmadani, M., and Pavlidis, P. (2018). Assessing Transcriptome Quality in Patch-Seq Datasets. Front. Mol. Neurosci. *11*, 363.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035–1041.

Vue, T.Y., Kim, E.J., Parras, C.M., Guillemot, F., and Johnson, J.E. (2014). Ascl1 controls the number and distribution of astrocytes and oligodendrocytes in the gray matter and white matter of the spinal cord. Development *141*, 3721–3731.

Wang, L.H., and Baker, N.E. (2015). E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. Dev. Cell *35*, 269–280.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635.

Weng, Q., Wang, J., Wang, J., He, D., Cheng, Z., Zhang, F., Verma, R., Xu, L., Dong, X., Liao, Y., et al. (2019). Single-Cell Transcriptomics Uncovers Glial Progenitor Diversity and Cell Fate Determinants during Development and Gliomagenesis. Cell Stem Cell *24*, 707–723.e8.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

Winterer, J., Lukacsovich, D., Que, L., Sartori, A.M., Luo, W., and Földy, C. (2019). Single-cell RNA-Seq characterization of anatomically identified OLM interneurons in different transgenic mouse lines. Eur. J. Neurosci. *50*, 3750–3771.

Yang, N., Ng, Y.H., Pang, Z.P., Südhof, T.C., and Wernig, M. (2011). Induced neuronal cells: how to make and define a neuron. Cell Stem Cell *9*, 517–525.

Zemankovics, R., Káli, S., Paulsen, O., Freund, T.F., and Hájos, N. (2010). Differences in subthreshold resonance of hippocampal pyramidal cells and interneurons: the role of h-current and passive membrane characteristics. J. Physiol. *588*, 2109–2132.

Zhang, H.M., Liu, T., Liu, C.J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y., and Guo, A.Y. (2015). AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Res. *43*, D76–D81.

Zhou, H., Su, J., Hu, X., Zhou, C., Li, H., Chen, Z., Xiao, Q., Wang, B., Wu, W., Sun, Y., et al. (2020). Glia-to-Neuron Conversion by CRISPR-CasRx Alleviates Symptoms of Neurological Disease in Mice. Cell *181*, 590–603.e16.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-β-III-Tubulin	Sigma-Aldrich	Cat# T8660; RRID: AB_477590
Mouse anti-GFAP	Dako	Cat# Z0334; RRID: AB_100013482
Rabbit anti-GFAP	Sigma-Aldrich	Cat# G3893; RRID: AB_477010
Rat anti-RFP	Chromotek	Cat# 5F8; RRID: AB_2336064
Rat anti-RFP	Rockland	Cat# 600-401-379; RRID:AB_2209751
Anti-MAP2	Millipore	Cat# MAB378, RRID:AB_94967
Anti-MAP	Millipore	Cat# AB5622, RRID:AB_91939
Anti-vGlut2	Synaptic Systems	Cat# 135402; RRID:AB_2187539
Anti-Gad65/67	Sigma-Aldrich	Cat# G5163; RRID:AB_477019
Anti-synaptophysin	Synaptic Systems	Cat# 101 011 RRID:AB_887824)
Anti-NeuN	Merck/Millipore	MAB377; RRID:AB_2298772
Anti-Ascl1	BD PharMingen	RRID:AB_396479
Anti-Neurog2	Gift from DJ Anderson, Caltech, California	N/A
Anti-Sox9	Sigma-Aldrich	Cat# AB5535; RRID:AB_2239761
Anti-Olig2	Millipore	Cat# MABN50; RRID:AB_10807410
Anti-Iba1	Synaptic System	Cat# 234013; RRID:AB_2661873
Anti-Dcx	Millipore	Cat# ab2253; RRID:AB_1586992
Anti-Mouse Alexa Fluor 488	Molecular Probes	Cat# A-21202; RRID: AB_141607
Anti-Mouse IgG2b 633	Innovative Research	Cat# A21146; RRID: AB_1500899
Anti-Mouse IgG1 647	Molecular Probes	Cat# A21240; RRID: AB_141658
Anti-Mouse IgG1 Biotin	Southernbiotech	Cat# 1070-08; RRID: AB_2794413
Anti-Rabbit Alexa Fluor 488	Molecular Probes	Cat# A21206; RRID: AB_141708
Anti-rabbit-Cy5	ImmunoResearch	Cat # 111-175-144 RRID: AB_2338013
Anti-rabbit Cy3	ImmunoResearch	Cat# 711-165-152 RRID: AB_2307443
Anti-Rat Cy3	ImmunoResearch	Cat# 112-165-167 RRID: AB_2338251
Streptavidin Alex Fluor 405	Thermo Fisher	Cat# S32351
Chemicals, peptides, and recombinant proteins		
EGF	GIBCO	Cat# PHG0311
bFGF	GIBCO	Cat# 13256029
Poly-D-Lysine	Sigma-Aldrich	Cat# P0899
B27	GIBCO	Cat# 17504044
HBSS medium	Thermo Fisher	Cat# 24020117
HEPES	Thermo Fisher	Cat# 15630080
DMEM/F12	Thermo Fisher	Cat# 10565018
trypsin/EDTA 0,25%	Thermo Fisher	Cat# 25200056
Neurobasal Medium	GIBCO	Cat# 21103149
Glucose	GIBCO	Cat# A2494001
GluataMAX	GIBCO	Cat# 35050061
OptiMEM – GlutaMAX	Thermo Fisher	Cat# 51985-026
EGTA	Sigma-Aldrich	Cat# E3889
Lipofectamine 2000	Thermo Fisher	Cat# 11668019
Triton X-100	Sigma-Aldrich	Cat# T9284
BDNF	Peprotech	Cat# 450-02
GDNF	Peprotech	Cat# 450-10

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
cAMP	Sigma-Aldrich	Cat# D0260
NT3	Peprotech	Cat# 450-03
N2	Invitrogen	Cat# 17502048
Forskolin	Sigma-Aldrich	Cat# F-6886
Dorsomorphin	Sigma-Aldrich	Cat# P-5499
Hydroxyl-Tamoxifen	Sigma-Aldrich	Cat# H-7904
Bovine Serum Albumine (BSA)	Sigma-Aldrich	Cat# A9418
Critical commercial assays		
Arcturus PicoPure RNA Isolation Kit	Thermo Fisher	Cat# 12204-01
Agencourt AMPure XP	Beckman Coulter	Cat# 10136224
Deposited data		
Bulk RNA-seq	This study	GEO: GSE174238
Patch-Seq – Figure 4	This study	GEO: GSE173977
Patch-Seq – Figure 5	This study	GEO: GSE173978
SC_vs_GM Astrocytes	This study	GEO: GSE173979
Experimental models: Organisms/strains		
C57BL/6	LMU animal Facility	N/A
Recombinant DNA		
RV CAG-Neurog2-IRES-DsRedExpress2	Gascón et al., 2016	N/A
RV CAG-Ascl1-IRES-DsRed	Heinrich et al., 2010	N/A
RV CAG-DsRedExpress2	Heinrich et al., 2010	N/A
RV CAG-Ascl1ERT2-IRES-DsRed	Masserdotti et al., 2015	N/A
RV CAG-ERT2-Neurog2-IRES-DsRed	This study	N/A
Software and algorithms		
ZEN software	Zeiss	https://www.zeiss.com/microscopy/en_us/products/microscope-software RRID:SCR_013672
ImageJ	ImageJ	https://imagej.net RRID: SCR_003070
GraphPad Prism 5.0	GraphPad Software	http://www.graphpad.com:443/ RRID:SCR_002798
Adobe Photoshop	Adobe Photoshop	https://www.adobe.com RRID: SCR_014199
Adobe Illustrator	Adobe Illustrator	https://www.adobe.com RRID:SCR_010279
Microsoft Excel	Microsoft Excel	https://www.microsoft.com/en-gb/ RRID:SCR_016137
RStudio	RStudio	https://www.rstudio.com/ RRID: SCR_000432
TopGO v.2.34.0	Alexa and Rahnenfuhrer, 2018	https://bioconductor.org/packages/release/bioc/html/topGO.html
DESeq2 v. 1.22.2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Ggplot2 v.3.2.0	Wickham, 2016	https://cran.r-project.org/web/packages/ggplot2/index.html
fgsea	Sergushichev, 2016	https://bioconductor.org/packages/release/bioc/html/fgsea.html
Other		
Aqua Poly/Mount	Polysciences	Cat# 18606-20
pluriStrainer Mini	pluriselect	Cat# 43-10040-40
NucBlue	ThermoFisher	Cat# R37605

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Giacomo Masserdotti (giacomo.masserdotti@helmholtz-muenchen.de).

Materials availability

Plasmids generated in this study are available upon request.

Data and code availability

- RNASeq data have been deposited to GEO (GSE174238; GSE173977; GSE173978; GSE173979) and publicly available. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Wild-type mice (primary cell culture, IHC, RNA-sequencing)

All experimental procedures in this study, done at the LMU Munich, were performed in accordance with German and European Union guidelines and were approved by the government of Upper Bavaria. Primary cultures of astrocytes from spinal cords were obtained from brains of C57BL/6J mice of P2-3 days of age; primary cultures of gray matter cortex were obtained from brains of C57BL/6J mice of P6-7 days of age; no specific gender was considered. Mice were fed *ad libitum* and housed with 12/12 h light and dark cycle and kept under specific-pathogen-free (SPF) conditions.

Primary cultures of astrocytes

Three to six postnatal (P2-P3) mice were sacrificed and their entire spinal cord isolated from the vertebrae, after carefully ablating the dorsal root ganglia. For cortical gray matter astrocytes, 2 postnatal (P6-P7) mice were sacrificed, brain extracted and only the gray matter of the cerebral cortex was isolated, after paying attention to remove the sub-ventricular zone and the hippocampus. Both spinal cords and gray matter tissue was dissociated to obtain a single cell suspension using the OctoMACs protocol (Miltenyi Biotec), according to manufacturer's instruction. Astrocytes were subsequently isolated using anti- ACSA-2 MACS-microBead Technology (Miltenyi Biotec) according to manual's instruction (Kantzer et al., 2017). Primary cultures of gray matter-derived astrocytes were expanded in uncoated plastic flasks, while spinal cord-astrocytes plated on poly-D-lysine (PDL, Sigma-Aldrich)-coated flasks. Cells were grown in T12 (spinal cord) or T25 (gray matter) flasks in medium consisting of DMEM/F12 (1:1) with GlutaMax, 10% fetal bovine serum (FBS), glucose, penicillin /streptomycin, and 1x B27 serum-free-supplement, 10 ng/ml epidermal growth factor (EGF), and 10 ng/ml basic fibroblast growth factor (bFGF) (astro-medium). Primary cultures of astrocytes were maintained in an incubator for 6-8 days at 37°C and 5% CO_2. Cells were passaged at 80%–90% confluency using 0.05% trypsin/EDTA and plated on poly-D-lysine coated glass coverslips at a density of 50,000-60,000 cells per coverslip (in 24-well plates) or in PDL-coated 6-well plates (Nunc) at a density of 300,000 cells per dish in fresh and complete astro-medium, supplemented with EGF and bFgf. Cells were either fixed at the time of infection, to evaluate the purity of the cultures, or infected with retroviral particles 12-16 hours after plating. For the experiment in Figure 6, cells were collected 1 day after re-plating.

METHODS DETAILS

Plasmids and viral production

The plasmid containing Ascl1ERT2 has been previously described (Masserdotti et al., 2015). Neurog2ERT2 was generated by cloning ERT2 at the 5′- of Neurog2 cDNA. The cDNA was cloned downstream of the CAG promoter and followed by an Intra-Ribosome-Entry-Site (IRES) and DsRedExpress2; similarly, Ascl1 and Neurog2 have been cloned in the same retroviral backbone, characterized by the CAG promoter (CMV enhancer, chicken beta-actin promoter and a large sysnthetic intron), and the IRES-DsRedExpress2 sequence; control virus was generated by cloning DsRedExpress2 downstream of the CAG promoter (Gascón et al., 2016; Heinrich et al., 2010, 2011). This allows to identify transduced cells and quantify the reprogramming efficiency over the transduced cells. To produce viral particles, retrovirus-encoding plasmids were transfected in 293GPG (Ory et al., 1996) and collected via ultracentrifugation at 27.000 rpm for 2 hours after 3, 5 and 7 days, as previously described (Heinrich et al., 2011). Pellet was resuspended in 100μl of PBS 1X (added with 5mM MgCl2) and aliquots stored at −80°C until use.

Transduction

Primary cultures of astrocytes were plated in 24-well plates at a density of 50,000-60,000 cells per well; the next day, cells were infected with viral particles, according to the experimental design. Twenty-four to 36 hours later, astro-medium was replaced with fresh

medium consisting of DMEM/F12 (1:1), penicillin/streptomycin, supplemented with 1x B27 and GlutaMax, glucose but not FBS, EGF and FGF (differentiation medium) and the cells maintained in culture until 6-7 days post-transfection in 9% CO_2 incubator. For FACSorting, astrocytes were plated in 6-well plates pre-coated with PDL at a concentration of 300,000-350,000 cells per well. The following day, cells were transduced with the retroviral particles. One day later, medium was replaced with differentiation medium and cells were harvested 24 hours later (see scheme in Figure 2A). The viruses used are listed in the key resource table and were produced as previously described (Heinrich et al., 2011; Masserdotti et al., 2015).

Long term culture of reprogrammed astrocytes

To improve the survival of the iNeurons for the characterization of the neuronal subtypes, electrophysiology, single-cell experiments (patch-seq), cultures were treated with maturation medium (BDNF 20ng/ml, GDNF 20ng/ml, N2, NT3 20ng/ml, cAMP 100μM) every fourth day, starting from day 8DPI. Catalog numbers are indicated in Key resource table.

Fluorescence-activated cell sorting

Astrocytes were washed twice with 1x PBS, treated with trypsin (0,05% in EDTA) for 5 minutes, before resuspending in pre-warmed DMEM/F12 phenol-red free medium was added. Cells were harvested by centrifugation (1,000 rpm, 5min, 4°C), washed twice with 1x PBS and resuspended in 300μl of DMEM/F12 (1:1), phenol-red-free. Single cell suspension was filtrated using a 40-μm cell strainer and a drop of NucBlue was added to label living cells. Gates were defined using negative control (un-transduced cultures of astrocytes) and positive (cultures of astrocytes transduced with DsRed-encoding virus). The following lasers were used: x axis: 582/15 (DsRed); y axis 530/30 (FITC) (no signal in the channel; detecting autofluoresence). 20000-to-30000 events were collected per sample in a 1.5 mL tube containing 300μl of 0.05% BSA in PBS1X. Samples were kept in ice until the end of sorting, then harvested by centrifugation (1,000 rpm, 5min, 4°C). RNA was extracted with Arcturus PicoPure RNA isolation kit, according to manufacturer's instruction.

Immunocytochemistry

Cells were fixed in 4% paraformaldehyde (PFA) in 1X PBS for 10 min. at room temperature, washed in 1X PBS twice for 5 minutes, and stored up to a month at 4°C before staining. Specimen were incubated in primary antibodies (see Key Resource table) in PBS1X containing 4% Bovine Serum Albumin (BSA) and 0.5% Triton X-100 for 2 hours at room temperature or overnight at 4°C. After washing three times for 5 minutes with PBS, cells were incubated with the appropriate species- or subclass-specific secondary antibodies, with or without DAPI to label nuclei (blue), diluted 1:10000, for 1 hour in the dark at room temperature. Optionally, after incubating with primary antibodies and washing with PBS, biotin-labeled secondary antibodies were used at a dilution of 1:200 for 1 hour, followed by streptavidin-coupled fluorophores (1:500) for another hour. Coverslips were then mounted with Aqua Poly/Mount. List of primary and secondary antibodies can be found in the Key Resource table.

Bulk RNA sequencing

Transcriptional changes upon Ascl1ERT2 or Neurog2ERT2 activation

Cells from five biological replicates per condition (DsRed+OHT, Ascl1ERT2+OHT, NeurogERT2+OHT) and 2 replicates for DsRed-OHT-untreated condition were sorted and their RNA extracted with PicoPure Kit (Arcturus, Kit0204). One entire biological replicate (DsRed+OHT, Ascl1ERT2+OHT, NeurogERT2+OHT) was removed from the analysis because of the low reads in one of the samples (DsRed+OHT). GEO number GSE174238.

Transcriptome of acutely isolated astrocytes cultured astrocytes

Following MACS procedure, cells were collected for RNA extraction ("acute samples"). Alternatively, MACS-sorted astrocytes were grown in flasks as described. Upon 80% confluency, cells were re-seeded in 6-well plates with fresh astrocyte-medium, as for being transduced. One day later, cells were collected and processed for RNA extraction. Five to six independent biological samples were collected per condition (GM n = 5 both acute and culture; SC, n = 6 both acute and culture). GEO number: GSE173979.

RNA was isolated on column using the PicPure™ RNAextraction kit (Applied Biosystems) and RNA quality and concentration were evaluated with an Agilent BioAnalyzer 2100 (Agilent). All included samples had a RIN > 9. 3μg of RNA from each sample was used to generate the RNA-seq libraries using bulk-adapted mcSCRB-seq protocol (Bagnoli et al., 2018): cDNA was generated by oligo-dT primers containing well-specific (e.g., sample specific) barcodes and unique molecular identifiers (UMIs). Unincorporated barcode primers were digested using Exonuclease I (Thermo Fisher). cDNA was pre-amplified using KAPA HiFi HotStart polymerase (Roche) and pooled before Nextera libraries were constructed from 0.8 ng of pre-amplified cleaned up cDNA using Nextera XT Kit (Illumina). 3′ ends were enriched with a custom P5 primer (P5NEXTPT5, IDT) and libraries were size selected using 2% E6 Gel Agarose EX Gels (Life Technologies), cut out in the range of 300–800 bp, and extracted using the Monarch DNA Gel Extraction Kit (New England Biolabs) according to manufacturer's recommendations. Libraries were paired end sequenced on an Illumina HiSeq 1500 instrument. Sixteen bases were sequenced within the first read to obtain cellular and molecular barcodes, and 50 bases were sequenced in the second read into the cDNA fragment. An additional eight bases were sequenced to obtain the i7 barcode. On average, we sequence around 20 million read/sample. Gene-based transcripts counts were obtained by running the zUMI pipeline (Parekh et al., 2018) (version 0.0.2) using Ensembl annotation release 81.

RNA-seq analysis

The analysis was performed using R (3.5.3) and RStudio (version 1.2.1335). See Key Resources Table for packages used. As no significant difference was found between OHT-treated and OHT-untreated Ds-Red-transduced cells (unsupervised clustering, Figure S2F, Principal Component analysis, Figure S2G, no genes differentially expressed with padj < 0.01, Data S1), we compared proneural factor induced programs to DsRed-OHT treated samples. Gene Ontology enrichment analysis was performed using the package "TopGO" in RStudio: differentially expressed genes (Log2FC > 1, padj < 0.01 or Log2FC < −1, padj < 0.01) were provided as input, while the list of the genes with a pvalue were used as background. Top 20 GO, ranked on the basis of Exact Fisher score (< 0.01), were selected (see Tables S2–S9; S11; S14; S16–S19; S21–S23; S25–S32 in Data S1,S2, S3, S4, S5, and S6). GO terms were then ranked for the enrichment, obtained by diving the number of detected genes versus the number of expected genes, and top 5 were plotted. Gene Set Enrichment Analysis (Subramanian et al., 2005) was performed using the package "fgsea" in Rstudio (Sergushichev, 2016). "An algorithm for fast pre-ranked gene set enrichment analysis using cumulative statistic calculation").

To define the Astrocyte- and Neuron-Score, normalized counts per each gene were obtained from the package DSEq2 (Love et al., 2014). Genes associated with astrocytes were obtained from Tripathy et al. (2018) and Weng et al. (2019), summed and used as coordinate to plot in ggplot2. To define the TFs in Figures 6K and 6L, TFs were filtered from all differentially expressed genes SC-Astro versus GM-Astro (padj < 0.01).

Single cell RNA-sequencing patch-seq related to Figures 4 and S3

Three- to four-week-old single cells were collected using an electrophysiology setup following a modified protocol (Cadwell et al., 2017). The coverslips were kept for 1-1.5 hours in the chamber for cell picking with a single cell collected every 5-6 minutes. Cells to be aspirated were selected by random screening of the coverslip. Brightfield and fluorescence pictures were taken before and after aspiration to catalog each cell for morphometric analysis and to confirm the successful aspiration of the soma into the electrode. We used borosilicate glass Warner Capillary Glass Tubing with polished ends and the inside filament (Model No. G150F-4) to make the aspiration electrodes with the input resistance between 2-3 MΩ (tested with the Axon Instruments Axopatch 200B amplifier). The intracellular solution contained (in mM): 4 KCl, 2 NaCl, 0.2 EGTA, 135 K-Gluconate, 10 HEPES (K-salt), 4 ATP (Mg-salt), 0.5 GTP (Na-salt), 10 Phospho-creatin, pH = 7.3, 290 mOsm. This intracellular solution was supplemented with RNase Out (1:40, 40 U/ μl, Invitrogen) and glycogen (1:1000 from 20 mg/ml stock, Roche). Each electrode was backfilled with 1.3 μL of this supplemented intracellular solution just before aspiration. To collect the individual aspirated iNeurons, PCR tubes were filled with the following solution as described for the SmartSeq2 method by Picelli et al. (2014). We modified some volumes and concentrations as suggested in the patch-seq method by Cadwell et al. (2017), because it worked better for our experimental setting. Each PCR tube contained 1 μL of 0.6% Triton X-100 (vol/vol) in RNase-free water (sterile, disk filter, stored at 4°C) with 1:40 of RNase Out (Invitrogen), 1 μL of oligo-dT_{30}VN primers (10 μM stock, stored at −80°C), and 1 μL of dNTP (10 mM stock, stored at −20°C). The PCR tubes were kept on ice until the cell was ready to be collected. The PCR tubes were kept at −80°C until ready for the SmartSeq2 reverse transcription and cDNA amplification. Libraries were prepared with MicroPlex Library Preparation Kit v2 (Diagenode, C05010014). Cells were collected from 2 biologically independent reprogramming experiments. Sequencing reads were mapped to the GRCm38 reference genome using STAR software version 2.6.0a. TPM expression values based on ENSEMBL annotation version GRCm38.92 were calculated with RSEM (1.3.0). Data are deposited with the number GSE173977 in GEO.

Single cell-RNA-seq Patch-seq related to Figures 5 and S4

The culture coverslips were transferred to the recording chamber and were constantly perfused with fresh aCSF heated to 28°C using an in-line temperature controller (SH-27B combined with TC-324C, Warner Instruments Corp., Connecticut, USA). The aCSF consisted of (in mM) 125 NaCl, 2.5 KCl, 26 NaHCO3, 1.25 NaH2PO4, 2 CaCl2, 2 MgCl2, 11 D-glucose (pH at 28°C = 7.4, osmolarity = 305–315 mOsm, perfused with a mixture of 95% O_2 and 5% CO_2). The coverslips were visualized under an upright microscope (Axioskop FS, Zeiss, Oberkochen, Germany) equipped with differential interference contrast (DIC)-infrared optics and epifluorescence (filter set: Zeiss BP450-490, LP520). The fluorescence and infrared images were acquired with a CCD camera (Orca-ER, Hamamatsu, Shizouka, Japan). Single cells were visualized using a 40 × /0.74 N.A. water immersion objective (Olympus). The electrodes for whole-cell patch-clamp recordings were pulled from borosilicate glass capillaries (OD = 1.5 mm, ID = 0.86 mm, Warner Instruments Corp., Connecticut, USA) using a puller (Zeitz-Instruments, Martinsried, Germany). The intracellular solution consisted of (in mM) 126 K-gluconate, 4 KCl, 10 HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid], 10 Na2-phosphocreatine, 4 Mg-ATP, 0.3 Na2-GTP, 0.05 Alexa Fluor 488 (Life Technologies), adjusted to pH 7.4 at 28°C and 295–300 mOsm. The electrodes (resistance ~4 MΩ) were connected to the amplifier's headstage via a chlorided silver wire. A silver / silver chloride-pellet immersed into the aCSF solution in the chamber served as reference electrode. Somatic whole-cell recordings were made in the current clamp mode using an ELC 03 XS amplifier (npi electronics, Tamm, Germany). Bias and offset current were zeroed before giga seal formation.

Electrophysiological recordings were performed as described in the section Electrophysiology with the few differences: following membrane rupture, the cells were not clamped to −60 mV and resting membrane potentials higher than −50 mV were taken as one of the parameters showing immature reprogrammed cells. Determination of the membrane resistance was only performed by measurement of the amplitude of a voltage deviation induced by a small hyperpolarizing current pulse (1 s, 5 – 10 pA). The cells' ability to generate action potentials was tested by depolarizing current ramps (50 ms) from 0 – 100 pA. The AP amplitude was measured

as the difference between the spike threshold and the spike maximum, the AP half-width was determined as width in ms at spike half-height. Electrophysiological data analysis was performed using MATLAB 2020a (Mathworks, Massachusetts, USA).

Based on the electrophysiological data, recorded cells were divided into 4 categories: Ds-Red transduced astrocytes (control astrocytes); not firing, immature and firing iNeurons. Reprogrammed cells were considered not firing if the resting membrane potential was higher than −40 mV and/or the action potential not present; immature neurons were designated if the resting membrane potential was between −40 mV and −50 mV and the action potential amplitude lower than 40 mV; firing neurons were defined as having a resting membrane potential lower than −50 mV, a train of action potential, whose amplitude was higher than 40mV.

After the recording, the cell was aspirated using previously published protocols (Földy et al., 2016; Winterer et al., 2019). The content of the cell was aspirated by applying negative pressure to the capillary. The intracellular solution did not contain any ribonuclease inhibitors and very small amount of intracellular solution (> 2 μl) was used in the glass capillary to perform the electrophysiological recordings. After the extraction of the cell, the content of the capillary was expulsed into a single 0.2 mL PCR tube by applying positive pressure and breaking off the capillary tip. The PCR tube contained 3 μl of lysis buffer solution which consisted of 0.1% Triton X-100, 5 mM dNTP Mix, 2.5 μM oligo-dT30VN primers, and 1 U/μl RNase OUT. The sample was briefly spun down using a centrifuge and flash-frozen on dry ice. The subsequent reverse transcription (RT) and cDNA amplification was done according to the SmartSeq2 protocol (Picelli et al., 2014). The resulting cDNA fragments were analyzed using High Sensitivity DNA Kit (Agilent Bioanalyzer, Santa Clara, USA). After sequencing, raw reads were de-multiplexed on in-house high-performance-cluster (HPC) using Je (version 2.0.2). The raw sequencing reads were aligned to Ensembl GRCm38 mouse reference genome using STAR aligner (version 2.7.1) with the GeneCounts parameter on. Further processing of the data was performed in RStudio using the Bioconductor package DESeq2 (Love et al., 2014). Bubble plots showing the expression of the genes in different conditions were generated as follows: normalized gene counts per each gene were divided by the gene length (in Kb, obtained from Ensembl); then, such expression values were averaged within each subgroup, and then log2 transformed. Data are deposited with the number GSE173978 in GEO.

Single-cell morphometry

Pictures of individual single cells were subjected to a morphometric analysis. Using ImageJ software, several parameters were manually analyzed: number of primary processes, angles between the processes, the longitudinal axis of the cell soma, and the cross-section area of the cell soma. As n = 2 independent samples, we used nonparametric test for the statistical analysis. For the distribution of the angles between processes, we separated the measured angles into bins separated by 20 degrees and used non-linear regression of the fourth-order polynomial function to curve fit the frequency distribution. All the collected cells were analyzed, with the exception of 1 cell from Neurog2 (Ascl1 n = 20; Neurog2 = 21).

Ranking of transcription factors

Putative lead F (transcription or chromatin modification factor) (Zhang et al., 2015) was assessed by an approach adopted from Rackham et al. (2016). For each factor we established a sphere of influence of up to three level depth using gene-gene relations based on STRING database version 10.5 (Szklarczyk et al., 2017). We considered only relations with total scores > 300, where less than half of the value was attributed to text mining. Using measured difference in RNA-seq expression (data processed with DESeq2; Love et al., 2014) we calculated scores one for F and one for its underling network using following equations:

$$Score_F = |log(FC_F)|(-\log(AdjPval_F))$$

$$Score_{Network} = \sum_{g=1}^{n} \frac{|log(FC_g)|(-\log(AdjPval_g))}{Dist_g\ Pnd_g}$$

where: F – factor, FC – fold change, AdjPval – adjusted p value (both from DSeq2), g – gene, Dist – number of steps between g and F, Pnd – parent node degree, n - length of list of genes associated with F.

Subsequently all factors were ranked based on combined ranking of both scores and filtered for the normalized expression (TFs with norm. expression < 200 after induction were excluded for Figure 2, while for Figure 3 genes with norm. exp < 20 were excluded). To provide context to factors driving cell identity, we plotted a network using Cytoscape 3.6 (Su et al., 2014a) with edges width correlating to STRING interaction score. Only factors connected to other TFs were plotted.

Comparison of scRNA-seq with publicly available data

To find if our dataset holds some resemblance to any cell type, we used available data (Delile et al., 2019). We combined iNeuron single cell RNA-seq and public data to calculate average gene expression values for each cell type and PCs (Principal Component) of resulting dataset. To compare cell type similarity, we plotted 3rd and 4th component.

Electrophysiology

For electrophysiological recordings, coverslips were transferred to an organ bath mounted on the stage of an upright microscope (Axioscope FS, Zeiss, Göttingen, Germany). The cell cultures were perfused with a bathing solution consisting of (in mM): NaCl

150, KCl 3, $CaCl_2$ 3, $MgCl_2$ 2, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) 10, and D-glucose 10. The pH of the solution was adjusted to 7.4 (NaOH) and its osmolarity ranged between 309 to 313 mOsmol. The perfusion rate was set to 1.4 mL / min and recordings were performed at room temperature (23 – 24°C). The microscope was equipped with differential interference contrast (DIC) optics and with epifluorescence optics for green and red fluorescence (filter sets: Zeiss BP450-490, LP520, Zeiss BP546/12, lP590). Images were taken and displayed using a software-operated CCD microscope camera (ORCA R, Hamamatsu, Herrsching, Germany).The electrodes for whole cell patch-clamp recordings were fabricated from borosilicate glass capillaries (OD: 1.5 mm, ID: 0.86 mm, Hugo Sachs Elektronik-Harvard Apparatus, March-Hugstetten, Germany) and filled with a solution composed of (in mM): potassium gluconate 135, KCl 4, NaCl 2, ethylene glycol-bis(2-aminoethylether)-*N*,*N*,*N*′,*N*′-tetraacetic acid (EGTA) 0.2, HEPES (potassium salt) 10, adenosine-triphosphate (magnesium salt, ATP[Mg]) 4, sodium guanosine-triphosphate (NaGTP) 0.5, and phosphocreatine 10 (pH: 7.25 – 7.30, osmolarity: 288 – 291 mOsmol). The series resistance determined after establishment of the whole cell recording mode (9 – 17 MΩ) was compensated by 70 – 85%. The recorded signals were amplified (x10 or x20), filtered at 10 or 20 kHz (current clamp) and 3 kHz (voltage clamp), digitized at a sampling rate of 10 or 20 kHz and stored on a computer for offline analysis. Data acquisition and generation of command pulses was done by means of a CED 1401 Power 3 system in conjunction with Signal6 data acquisition software (Cambridge electronic design, Cambridge, England). Data analysis was performed using IGOR Pro 6 (WaveMetrics, Lake Oswego, USA) together with the NeuroMatic IGOR plugin (Version 2) (www.neuromatic.thinkrandom.com). Microscope images were corrected for contrast and brightness by using Photoshop CS3 (Adobe Software Systems, Ireland). By means of the motorized microscope stage, each coverslip was scanned systematically and iNeurons were identified by red fluorescence. Following membrane rupture, the cells were voltage-clamped to a holding potential of −60 mV and kept under this condition until stabilization of the holding current (3 – 5 min). Determination of the input resistance R_N was performed either by measurement of the amplitude of a voltage deviation induced by a small hyperpolarizing current pulse (1 s, 5 – 10 pA) or by determining the slope of the current-voltage-curve (IV-curve) at its origin. The somatic membrane time constant Ω was derived by fitting a dual exponential function to the voltage relaxation following cessation of a small hyperpolarizing current pulse and the total membrane capacity C_N was estimated using a method (Zemankovics et al., 2010). The cells' ability to generate action potentials was tested by injecting depolarizing current pulses (50 ms) with increasing current strengths (ΩI: 5 or 10 pA) or by depolarizing current ramps (50 ms) from 0 – 100 pA. The amplitudes of the action potentials (spikes) were measured as the difference between the resting membrane potential and the spike maximum, the spike duration was determined at half-maximum amplitude and the spike threshold was derived from a phase-plane plot (Bean, 2007). The action potential discharge pattern of the cells was investigated by injections of depolarizing current pulses (1 – 2 s), the amplitudes of which were raised in steps (5 or 10 pA) from 0 – 200 pA at a frequency of 0.1 Hz.

All chemicals and drugs were obtained from Sigma-Aldrich (Munich, Germany) and Biotrend (Cologne, Germany), respectively. Data are given as median ± IQR. Statistical comparison of two samples was performed by using Mann-Whitney unpaired two-tailed t test.

QUANTIFICATION AND STATISTICAL ANALYSIS

Immunostainings were analyzed with an AxioM2 epifluorescence microscope or LSM710 laser-scanning confocal (Carl Zeiss). Digital images were captured using the ZEN2 software (Carl Zeiss). Retroviral vector-transduced cells were quantified from more than 25 randomly chosen 20x fields in at least 3 independent experiments. Quantification for neuronal cells was based on β-III-tubulin immunoreactivity and morphological parameters, e.g., presence of 2 or more processes longer than 3x the cell soma as in Gascón et al. (2016). Astrocytes were quantified based on morphological features and expression of known astrocytic markers (Gfap or Sox9). Analysis of the culture composition (Figure S1) was performed by quantifying 2 tile images, each composed of 25 images (acquired at 20X) in 3 independent biological experiment. To quantify the intensity of Asc12ERT2 and Neurog2ERT2 (Figure S2B), 1 tile image (25 images at 20X) was imported in Fiji: nuclei were identified via DAPI and used to create a mask (define threshold; Set Measurement; Analyze particles (60 < particles < 250). Then, the mask was applied to the tile image to obtain the average intensity per each channel per each particle (that is, each nucleus present in the tile image). We selected the cells whose intensity value of green channel (corresponding to the transcription factor) was clearly above green-negative nuclei. Each tile image provided at the intensity of at least 170 cells. Data were collected for 3 independent cultures and analyzed in RStudio. Average intensity was calculated per each independent experiment. Boxplots show the median and the interquantile range (IQR) Whiskers are calculated as 1.58*IQR (Figures 1S–1U, S1H, S1K–S1P, S2D, and S4A–S4D). Data presented as bar plots show mean and confidence interval (CI) (Figures 1C, 1D, 2C, 3F–3I, and S3B–S3D).

Data were analyzed with Microsoft Excel, GraphPad Prism 7.0 software RStudio (version 1.2.1335); statistics was performed with linear regression using "lm" function (R Stats package) in RStudio on log2 transformed reprogramming efficiency. Evaluation of the residuals for fitted linear models was performed with the package "DHARMa" (Hartig, 2021). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.3.2.0. https://cran.r-project.org/web/packages/DHARMa/index) in RStudio. Statistical details of the experiments can be found in the figure legends. Significance is based on the p value indicated on the graphs as * $p \leq 0.05$, ** $p \leq 0.01$, ***$p \leq 0.001$.

Cell Reports, Volume *36*

Supplemental information

Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2

J. Kempf, K. Knelles, B.A. Hersbach, D. Petrik, T. Riedemann, V. Bednarova, A. Janjic, T. Simon-Ebert, W. Enard, P. Smialowski, M. Götz, and G. Masserdotti

Cell Reports, Volume *36*

Supplemental information

Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2

J. Kempf, K. Knelles, B.A. Hersbach, D. Petrik, T. Riedemann, V. Bednarova, A. Janjic, T. Simon-Ebert, W. Enard, P. Smialowski, M. Götz, and G. Masserdotti

Figure S1

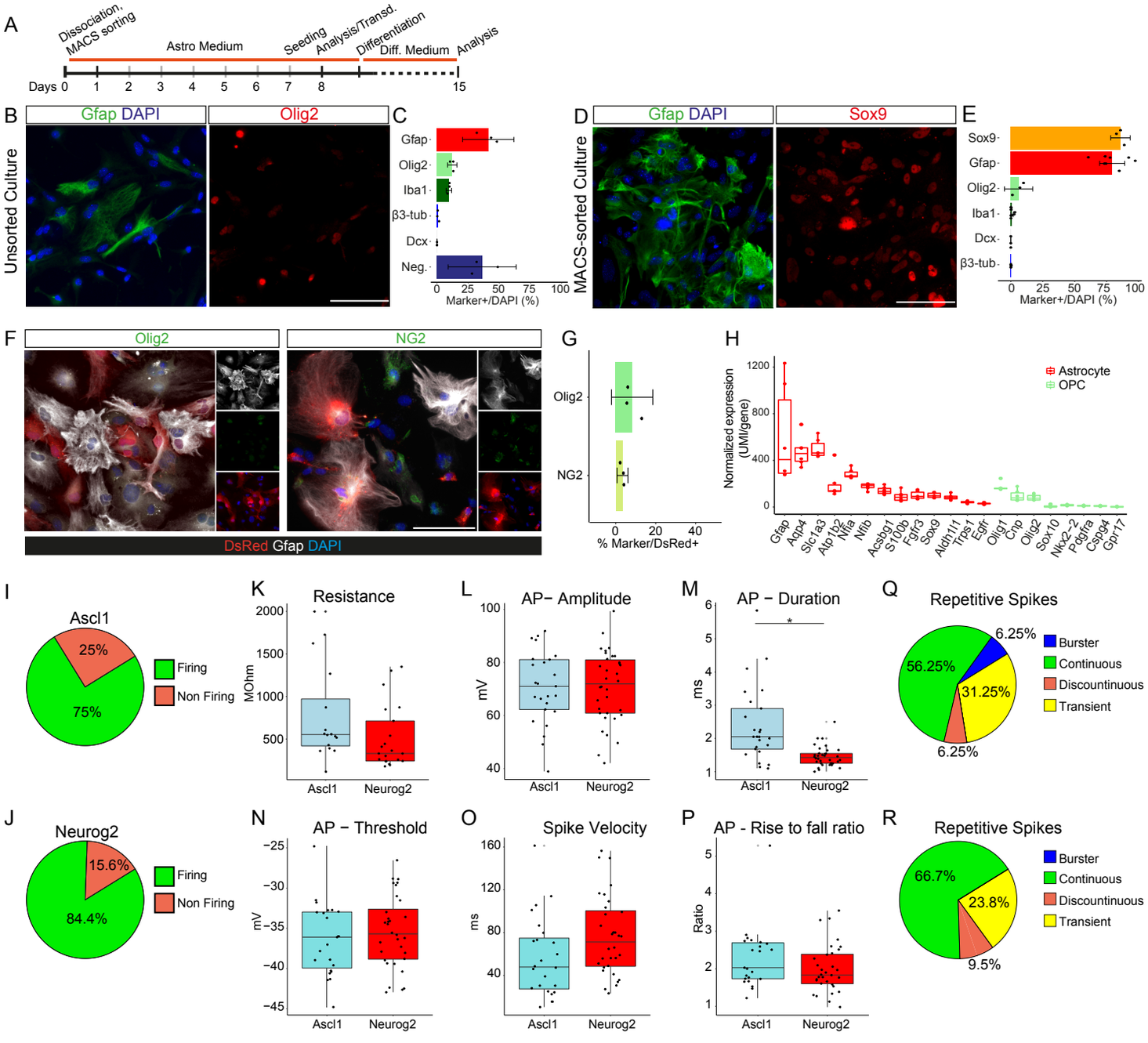

A
Dissociation, MACS sorting
Astro Medium
Seeding
Analysis/Transd.
Differentiation
Diff. Medium
Analysis
Days 0 1 2 3 4 5 6 7 8 15
B
Gfap DAPI
Olig2
Unsorted Culture
C
Gfap
Olig2
Iba1
β3-tub
Dcx
Neg.
0 25 50 75 100
Marker+/DAPI (%)
D
Gfap DAPI
Sox9
MACS-sorted Culture
E
Sox9
Gfap
Olig2
Iba1
Dcx
β3-tub
0 25 50 75 100
Marker+/DAPI (%)
F
Olig2
NG2
DsRed Gfap DAPI
G
Olig2
NG2
0 20 40
% Marker/DsRed+
H
Normalized expression (UMI/gene)
1200
800
400
0
Astrocyte
OPC
Gfap Aqp4 Slc1a3 Atp1b2 Nfia Nfib Acsbg1 S100b Fgfr3 Sox9 Aldh1l1 Trps1 Egfr Olig1 Cnp Olig2 Sox10 Nkx2-2 Pdgfra Cspg4 Gpr17
I
Ascl1
25%
75%
Firing
Non Firing
K
Resistance
MOhm
2000
1500
1000
500
Ascl1 Neurog2
L
AP− Amplitude
mV
100
80
60
40
Ascl1 Neurog2
M
AP − Duration
ms
*
6
5
4
3
2
1
Ascl1 Neurog2
Q
Repetitive Spikes
6.25%
56.25%
31.25%
6.25%
Burster
Continuous
Discountinuous
Transient
J
Neurog2
15.6%
84.4%
Firing
Non Firing
N
AP − Threshold
mV
−25
−30
−35
−40
−45
Ascl1 Neurog2
O
Spike Velocity
ms
160
120
80
40
Ascl1 Neurog2
P
AP - Rise to fall ratio
Ratio
5
4
3
2
1
Ascl1 Neurog2
R
Repetitive Spikes
66.7%
23.8%
9.5%
Burster
Continuous
Discountinuous
Transient

Figure S1 – Enrichment of astrocytes by ACSA2-based MACS-sorting, related to Figure 1

A) Schematic representation of the experimental design.

B, D) Micrographs of cultures after 7 days *in vitro* obtained from unsorted (B) or ACSA2-MACS-sorted (D) cells from P2 spinal cord. Scalebar = 100μm.

C, E) Bar plots showing the percentage of cells immune-positive for the indicated markers in the respective cultures (from unsorted, C, or sorted, E, cells). n=3 (C), n=3-6 (E) per each marker.

F, G) Micrograph (F) and bar plot (G) showing oligodendrocyte progenitor cells (OPCs) labeled by immunostaining for Olig2+ or NG2+ at the end of the reprogramming protocol (day 15).

H) Bar plot showing the normalized expression of astrocyte (red) and OPCs genes (green) following 7 days in culture (see Figure S6 for full transcriptome analysis).

I, J) Pie charts depicting the proportion of Ascl1-iNs (I) or Neurog2-iNs (J) showing an action potential.

K-P) Graphs showing the resistance of Ascl1- (n=15 cells) and Neurog2- (n=21 cells) iNs (K), the amplitude (L, Ascl1 = 25 cells; Neurog2 n=34 cells), the duration (M) (Ascl1 = 25 cells; Neurog2 n=34 cells), the threshold (N, Ascl1 = 23 cells; Neurog2 n=32 cells) of action potentials, the spike velocity (O, Ascl1 = 24 cells; Neurog2 n=32 cells) and the rise-to-fall ratio (P, Ascl1 = 23 cells; Neurog2 n=32 cells) at 28-42 DPI. Each dot represents a cell; n=5 independent experiments. *$p < 0.05$

Q, R) Pie chart depicting the different types of spike discharge patterns evoked in Ascl1-iNs (Q) or Neurog2-iNs (R).

Figure S2

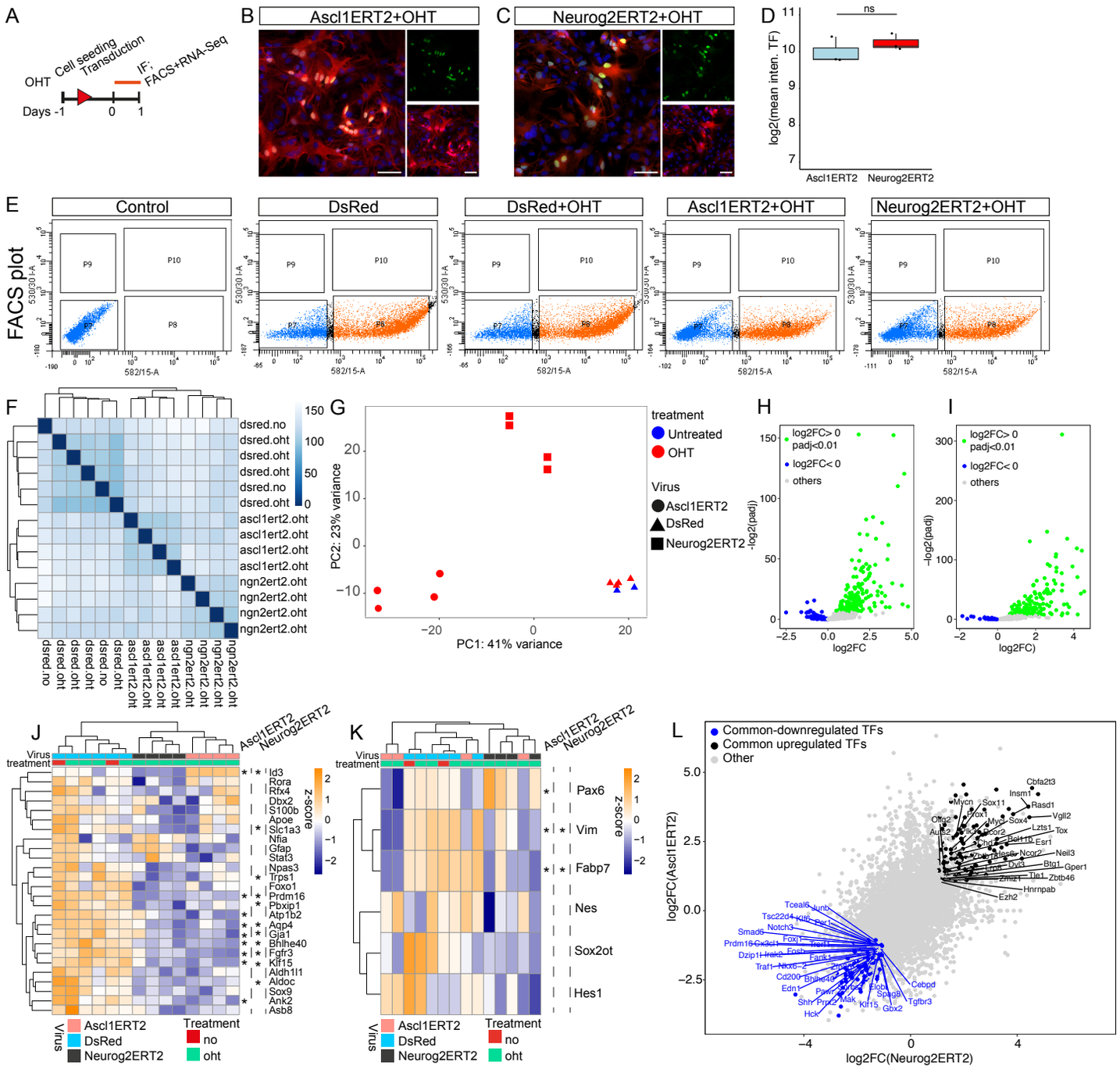

A
OHT
Cell seeding
Transduction
IF;
FACS+RNA-Seq
Days -1
0
1
B
Ascl1ERT2+OHT
C
Neurog2ERT2+OHT
D
ns
log2(mean inten. TF)
Ascl1ERT2
Neurog2ERT2
E
FACS plot
Control
DsRed
DsRed+OHT
Ascl1ERT2+OHT
Neurog2ERT2+OHT
P9
P10
P7
P8
582/15-A
F
dsred.no
dsred.oht
ascl1ert2.oht
ngn2ert2.oht
G
PC2: 23% variance
PC1: 41% variance
treatment
Untreated
OHT
Virus
Ascl1ERT2
DsRed
Neurog2ERT2
H
log2FC> 0
padj<0.01
log2FC< 0
others
-log2(padj)
log2FC
I
log2FC> 0
padj<0.01
log2FC< 0
others
-log2(padj)
log2FC)
J
Ascl1ERT2
Neurog2ERT2
Virus
treatment
Id3
Rora
Rfx4
Dbx2
S100b
Apoe
Slc1a3
Nfia
Gfap
Stat3
Npas3
Trps1
Foxo1
Prdm16
Pbxip1
Atp1b2
Aqp4
Gja1
Bhlhe40
Fgfr3
Klf15
Aldh1l1
Aldoc
Sox9
Ank2
Asb8
z-score
Virus
Ascl1ERT2
DsRed
Neurog2ERT2
Treatment
no
oht
K
Pax6
Vim
Fabp7
Nes
Sox2ot
Hes1
L
Common-downregulated TFs
Common upregulated TFs
Other
log2FC(Ascl1ERT2)
log2FC(Neurog2ERT2)
Cbfa2t3
Insm1
Mycn
Sox11
Rasd1
Fox1
Vgll2
Sox4
Lzts1
Tox
Esr1
Neil3
Btg1
Gper1
Tle1
Zbtb46
Hnrnpab
Ezh2
Tceal6
Junb
Tsc22d4
Klf6
Per1
Smad6
Notch3
Prdm16
Cx3cl1
Fosl1
Dzip1
Irak2
Fank1
Traf1
Nkx6-2
Cd200
Bhlhe40
Edn1
Pax6
Sox9
Elob
Spag8
Cebpd
Shh
Prox1
Mak
Klf15
Tgfbr3
Gbx2
Hck

Figure S2 – Molecular analysis of early (24 hours) transcriptome changes in SC-derived astrocyte direct neuronal reprogramming, related to Figure 2

A) Scheme of the experimental design for cell transduction, sorting and analysis.

B, C) Micrographs depicting the immunostaining for Ascl1 (B) and Neurog2 (C) showing the nuclear localization of Ascl1ERT2 and Neurog2ERT2 following 24 hours of treatment with OHT as quantified in (D). Scalebar = 50μm.

D) Boxplot showing the pixel intensity (log2(average intensity of FITC channel)) of nuclear Ascl1ERT2 and Neurog2ERT2 signal as shown in B and C following 24 hours of OHT treatment.

E) Examples of FACS gates used to sort DsRed positive cells. Control cells (un-transduced astrocytes from the same batch) were used to set the background auto-fluorescent gate.

F) Unsupervised clustering of sequenced samples.

G) Principal Component Analysis of the biological replicates. Note that DsRed-sorted un-treated (blue triangle) and OHT-treated DsRed control samples (red triangle) cluster together; conversely, Ascl1ERT2-OHT-treated samples (red squares) and Neurog2ERT2-OHT-treated samples (red circles) cluster apart from each other.

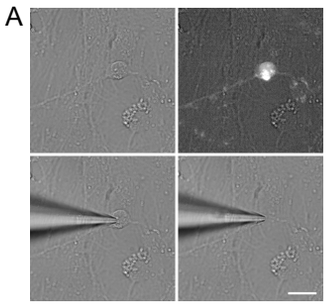
H, I) Volcano plots depicting the log2FC of genes upregulated by Ascl1ERT2 (Figure 2E) following Neurog2ERT2 activation (H) or log2FC of genes upregulated by Neurog2ERT2 following Ascl1ERT2 activation (I). Genes specifically upregulated by a given TF are those downregulated or unchanged by the TF (identified by blue dots in both graphs).

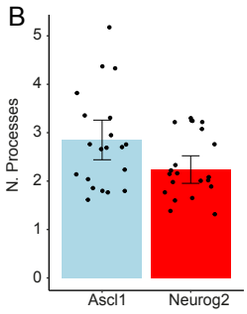
J) Heatmap of the expression of astrocytic markers (Weng et al., 2019) at 24 hours after OHT. Asterisk (*) indicates genes statistically significant (padj<0.01 with a normalized expression higher than 100).

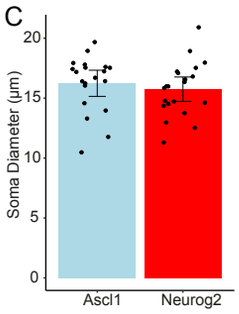
K) Heatmap of the expression of stem cell-related genes (Treutlein et al., 2016) at 24 hours after OHT. Asterisk (*) indicates genes statistically significant (pAdj<0.01 with a normalized expression higher than 100).

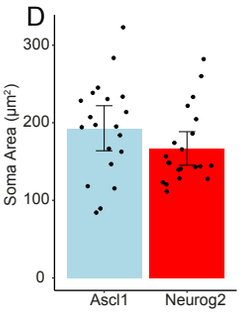
L) Volcano plot showing the genes differentially expressed between Neurgo2ERT2 and Ascl1ERT2 at 24 hours after OHT. Transcription factors up-regulated by both factors are indicated in black; downregulated by both in blue.

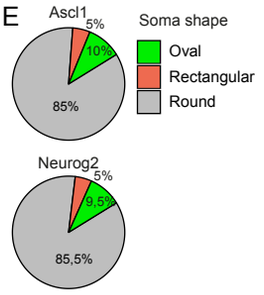
Figure S3











F

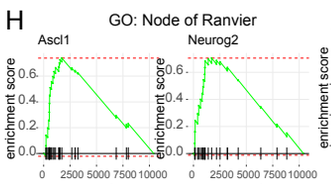
Ascl1-specific genes

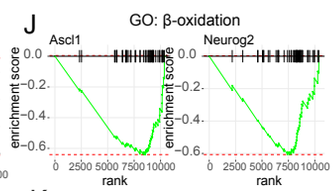
A830012C17Rik	Lhx3
Arc	Ngfr
Cyb561	Prdx5

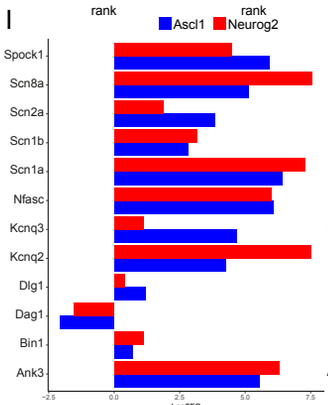
G

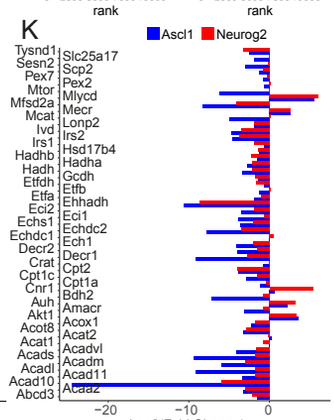
Neurog2-specific genes

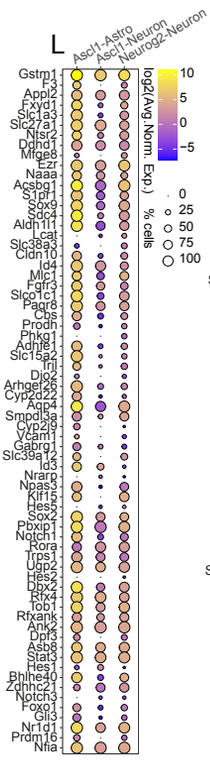
2300009A05Rik	Chst15	Neurog2
Alkbh4	Dach1	Synpr
Chn2	Mdga1	Zfp385b

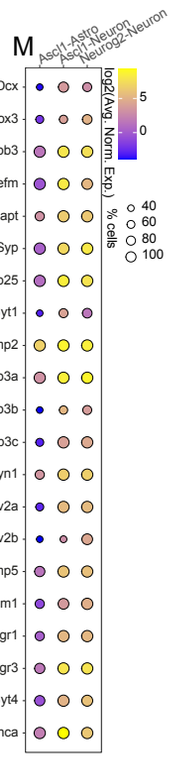


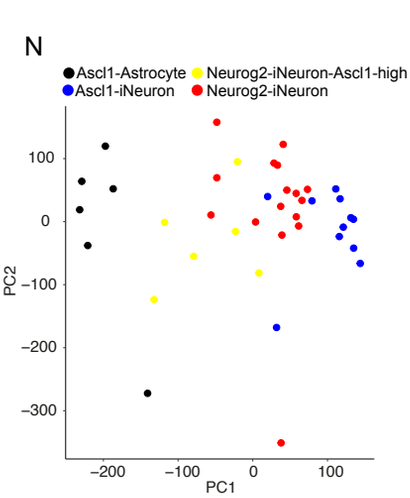


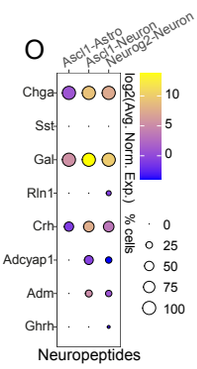


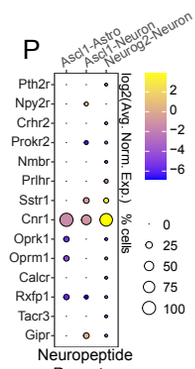


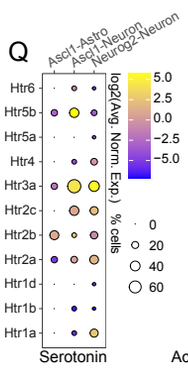












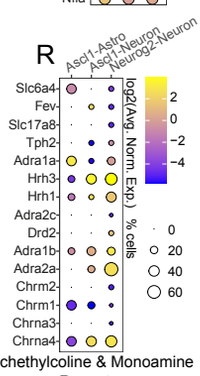


Figure S3 – Molecular analysis of Ascl1- and Neurog2-iNs at 3-4 weeks of reprogramming, related to Figure 4

A) Micrographs illustrating the collection of single cells for sequencing: a transduced (DsRed+) cell with neuronal morphology (upper panels) is approached with a capillary to establish a GigaOhm seal (low left panel); after aspirating the cell content, the cell is no longer visible (low right panel). Scalebar = 20μm.

B-E) Barplots (B-D) and pie diagrams (E) of the number of processes (B), the diameter (C), the area (D) and the shapes (E) of the somata of Ascl1- (n=20) and Neurog2- (n=21) iNs from 2 independent experiments on 21-28 DPI.

F, G) Genes specifically expressed by Ascl1-iNs (F) and Neurog2-iNs (G).

H) Example of terms from GSEA enriched in both Ascl1- and Neurog2-iNs.

I) Barplot depicting log2FC of the genes associated with the term “Node of Ranvier” in either Ascl1- (blue bars) or Neurog2-iNs (red bars).

J) Example of terms from GSEA enriched in control versus both Ascl1- and Neurog2-iNs.

K) Barplot depicting log2FC of the genes associated with the term “β-oxidation” in either Ascl1- (blue bars) or Neurog2-iNs (red bars).

L, M) Bubble plot showing the percentage of cells and log2(average expression) astrocyte (L) and neuronal markers (M) in not reprogrammed Ascl1-transduced cells (Ascl1-Astro), Ascl1-iNs and Neurog2-iNs. The expression of these genes was used to generate the astrocyte-score and neuronal-score (Figure 4H).

N) PCA showing the position of Neurog2-iNs with high expression level of endogenous *Ascl1* (in yellow) relative to other iNs.

O-R) Bubble plots showing the percentage of cells and log2(average expression) of neuropeptides (O), neuropeptide receptors (P), serotonin receptors (Q) and Achetylcholine and monoamine receptors (R) in not reprogrammed Ascl1-transduced cells (Ascl1-Astro), Ascl1-iNs and Neurog2-iNs collected at 21-28 DPI.

Figure S4

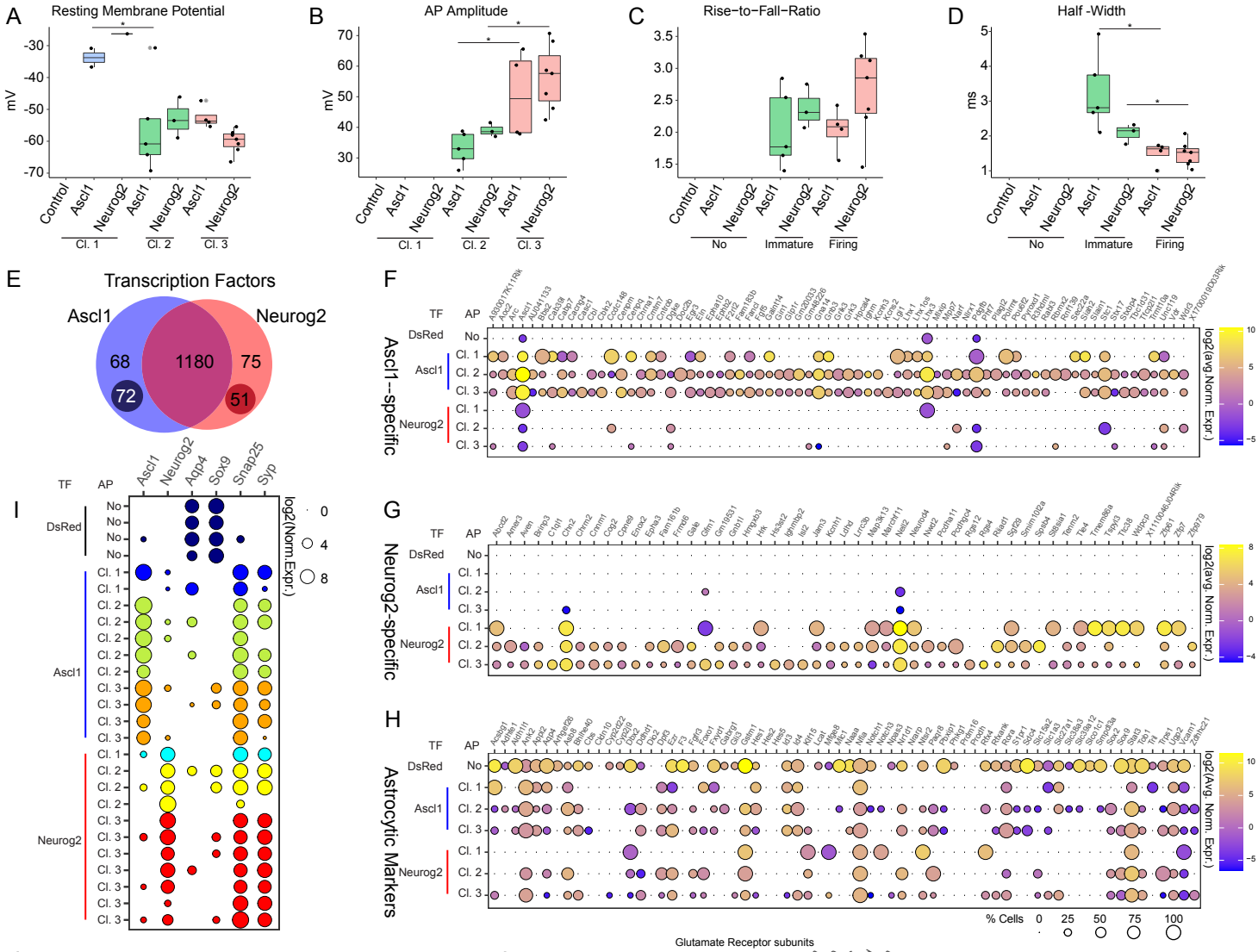

A
Resting Membrane Potential
mV
Control
Ascl1
Neurog2
Cl. 1
Cl. 2
Cl. 3
B
AP Amplitude
C
Rise−to−Fall−Ratio
No
Immature
Firing
D
Half -Width
ms
E
Transcription Factors
Ascl1
Neurog2
68
1180
75
72
51
F
Ascl1−specific
TF
AP
DsRed
log2(avg.Norm. Expr.)
G
Neurog2−specific
H
Astrocytic Markers
% Cells
I
Aqp4
Sox9
Snap25
Syp
log2(Norm.Expr.)
J
GO terms - Ascl1 specific genes
BP
positive regulation of glutamate receptor signaling pathway
regulation of keratinocyte proliferation
metanephric nephron development
positive regulation of signaling receptor activity
cell differentiation in spinal cord
CC
heterotrimeric G−protein complex
GTPase complex
plasma membrane receptor complex
receptor complex
plasma membrane protein complex
Enrichment
K
GO terms - Neurog2 specific genes
neuron cellular homeostasis
spinal cord motor neuron differentiation
cell differentiation in spinal cord
ventral spinal cord development
cellular response to retinoic acid
compact myelin
synaptic cleft
extrinsic component of postsynaptic membrane
axolemma
SAGA−type complex
L
Glutamate Receptor subunits
log2(Avr. Norm. Expr.)
%Cells
M
Glutamatergic markers
N
GABAergic markers
O
Glycinergic markers
P
PC4
PC3
Immature
Firing
No
Null_Progenitor
Null_Neuron
Not Firing
Control
Neuron
Progenitors
Neural Crest Progenitors
Neural Crest Neuron
DsRed−Control
Reprogrammed Neurons

Figure S4 – Patch-seq analysis of Ascl1- and Neurog2-iNs from day 14 onwards of reprogramming, related to Figure 5

A-D) Box plots showing the (A) resting membrane potential, (B) action potential amplitude, (C) rise-to-fall ratio and action potential half-width (D) of cells transduced by control virus or the factors indicated at 14-43 DPI. Each dot represents a cell; DsRed-Control, n=4; Ascl1 n=11 (Cl. 1 = 2; Cl. 2 = 5; Cl. 3 = 4); Neurog2, n= 11 (Cl. 1 = 1; Cl. 2 = 3; Cl. 3 = 7). N=3 independent experiments. *p<0.05

E) Venn diagram showing the number of genes upregulated (Log2FC>1, padj<0.01) in both Ascl1- and Neurog2-reprogrammed cells. In darker small circles, genes unique to a specific TF are highlighted.

F) Bubble plot depicting the average expression of 72 genes specific to Ascl1-iNs.

G) Bubble plot depicting the average expression of 51 genes specific to Neurog2-iNs.

H) Bubble plot showing the percentage of cells and log2(average expression) of astrocyte markers (from Weng et al., 2019 and Trypathy et al., 2018) in different subgroups.

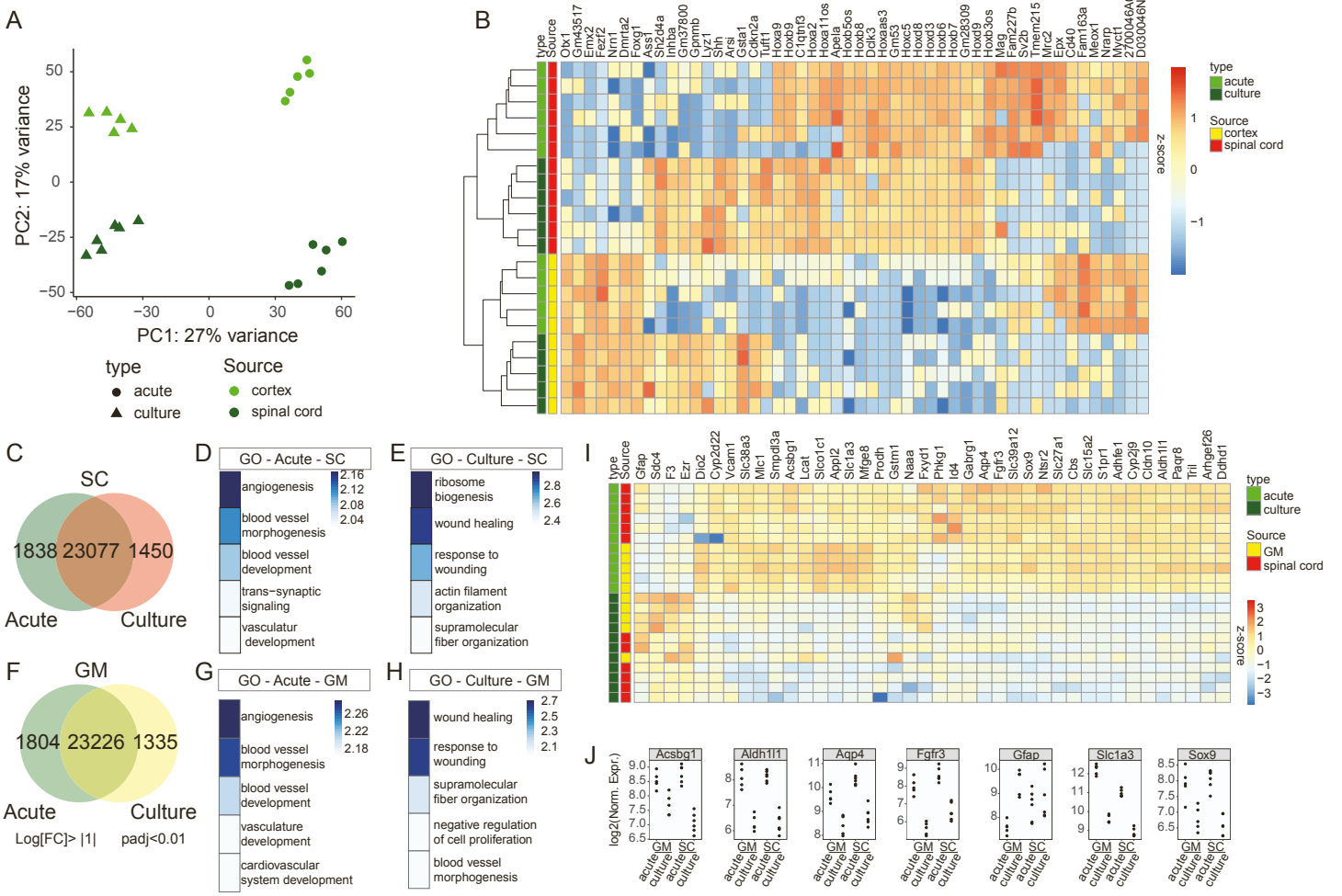
I) Bubble plot showing the expression of selected genes (proneural factors Ascl1 and Neurog2, the astrocyte markers Aqp4 and Sox9, and the synaptic markers Snap25 and Synaptophysin) in each cell analyzed.

J, K) Top 5 GO terms (BP, left; Cellular Compartment (CC), right) on 72 genes specifically expressed in Ascl1 reprogrammed cells (J, from Figure S4F) or in Neurog2 reprogrammed cells (K, from Figure S4G)

L-O) Bubble plots showing the percentage of cells and log2(average expression) of glutamate receptor subunits (L), Glutamatergic markers (M) and GABAergic markers (N) and glycinergic makers (O) in different subgroups of collected cells.

P) PC analysis of patch-seq cells (clustered as control, not firing, immature and firing) compared to SC progenitors and neurons from Delile et al. 2019.Note that DsRed-control cells are far apart from the endogenous neurons and reprogrammed cells.

Figure S5



K

2810468N07Rik	Cdk5r1	Dpp6	Gng2	Mast4	Pde1c	Rph3al	Tbc1d8
Actn3	Cdkn1c	Dpysl5	Gprin1	Mchr1	Pde2a	Rtn4rl1	Tbc1d9
Adcy5	Cdon	Dusp15	Grin2c	Megf11	Pdgfa	Rufy3	Tead2
Adm	Cerk	Ecel1	Grpr	Mex3b	Pkib	Scn5a	Tex14
Afap1	Chgb	Eepd1	Hdac11	Mfng	Plk3	Sema6b	Tmcc3
Aif1l	Chrna4	Elavl3	Hpcal1	Mkrn1	Plxnd1	Sftpc	Tmem191c
Als2cl	Cited4	Elavl4	Htr3a	Mn1	Pnmal2	Shisa2	Tmem201
Angptl2	Cldn3	Epha7	Id1	Mpzl2	Podxl2	Slc12a2	Tnnt2
Ar	Clip3	Ephb1	Id2	Mxra7	Polr3k	Slc26a2	Top1
Arc	Col18a1	Eya2	Iffo1	Myt1	Ppargc1b	Slc8a3	Tprn
Arpp21	Coro2b	Fam110a	Kcnip3	Nbl1	Ppp1r14b	Smoc1	Trp53i11
Atoh8	Ctrb1	Fam57b	Kcnq4	Ncald	Pvalb	Snca	Tspan14
BC034090	Cyth1	Fbxo32	Kcnt1	Ncor2	Pwwp2b	Socs2	Tspan18
Bcl11b	Dcx	Fbxw7	Klf10	Nefl	Ralgds	Sox4	Uck2
Bcl7a	Ddc	Filip1	Klhl21	Nfix	Rasd2	Sox8	Usp3
Bmp2	Dgkz	Flrt1	Kremen1	Ngfr	Rbm38	Spock1	Ust
Bmp7	Dleu7	Fxyd6	Lhx3	Nos1	Rgmb	Srpk3	Vash2
Brsk2	Dlgap4	Gal	Lmo1	Nptx1	Rgr	Ssbp3	Whamm
Cacna2d2	Dll1	Gdpd1	Lrrn1	Obsl1	Rhbdl3	Sstr2	Wnk2
Calm1	Dll3	Gga2	Mapk4	Olfm1	Rimbp3	Svil	Wnt9a
Ccnjl	Doc2b	Glce	Mapk8ip1	Olfml2b	Rnf114	Syn2	Zc3h12c
Cdh22	Dok7	Gna14	Mapkapk2	Otof	Rnf122	Syt13	Zdhhc14

L

4930579G24Rik
Adamts4
Cadm3
Chn2
Cmtm7
Cnr1
Cplx2
Dll4
Emid1
Fndc5
Kctd14
Lrrtm1
Ly6e
Mkx
Mrc2
Neurod4
Porcn
Ppp1r14a
Qsox1
Sema5b
Shb
Tmem2

Figure S5 – Molecular analysis spinal cord (SC)- and cerebral cortex Grey Matter (GM)-derived astrocytes, related to Figure 6

A) PC analysis on the GM-astrocytes (green dots) and SC-astrocytes (dark green) either acutely isolated (dots) or following 8 days in culture (triangle). n=5 GM-astrocytes, n=6 SC-astrocytes.

B) Heatmap showing the relative expression of the top 50 most variable genes across all the samples (acute GM-astrocytes, cultured GM-astrocytes, acute SC-astrocytes, cultured SC-astrocytes).

C, F) Venn diagram depicting the genes differentially expressed between acutely isolated and cultured astrocytes from the SC (C) or from GM (F).

D, E) Top 5 GO terms related to biological process from genes highly expressed in acutely isolated (D) or cultured (F) astrocytes from spinal cord.

G, H) Top 5 GO terms related to biological process from genes highly expressed in acutely isolated (G) or cultured (H) astrocytes from cortical GM.

I) Heatmap of astrocyte genes expressed in the 4 groups analyzed: note that the pan-astrocyte markers are expressed at different levels depending on the time of isolation (acute or after culturing for 7 days). n=5 GM-astrocytes, n=6 SC-astrocytes.

J) Dot plots showing the normalized expression of a subset of astrocytic markers in different conditions. Each dot represents an independent experiment. Note that cultured astrocytes have a lower expression of many astrocyte marker genes.

K) List of genes commonly upregulated upon Ascl1ERT2 activation for 24 hours in GM- and SC-derived astrocytes (log2FC>2, padj<0.01 for SC; log2FC>1 and pval<0.05 for GM).

L) List of genes commonly upregulated upon Neurog2ERT2 activation for 24 hours in GM- and SC-derived astrocytes (log2FC>2, padj<0.01 for SC; log2FC>1 and pval<0.05 for GM).

Cell Reports

CellPress
OPEN ACCESS

Correction

Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneural factors Ascl1 and Neurogenin2

J. Kempf, K. Knelles, B.A. Hersbach, D. Petrik, T. Riedemann, V. Bednarova, A. Janjic, T. Simon-Ebert, W. Enard, P. Smialowski, M. Götz,* and G. Masserdotti*

*Correspondence: magdalena.goetz@helmholtz-muenchen.de (M.G.), giacomo.masserdotti@helmholtz-muenchen.de (G.M.)

https://doi.org/10.1016/j.celrep.2021.109571

(Cell Reports *36*, 109409-1–109409-16.e1–e7; July 20, 2021)

In the originally published version of this article, Figure 3G contained an incorrect number of biological replicates. The original and corrected figure appear here, and the corrected figure now appears with the paper online. This correction does not change the percentages reported in the text.

The authors regret these errors.

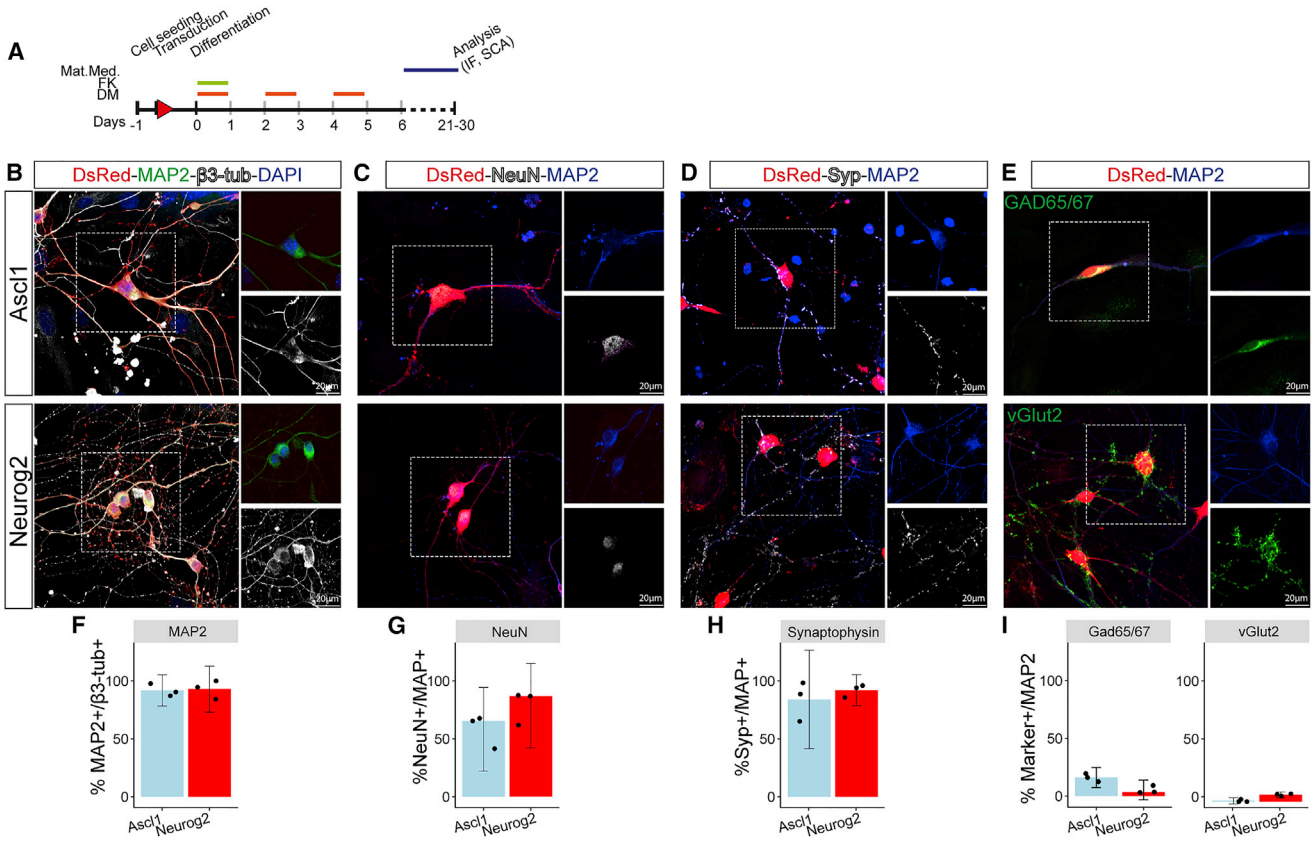


Figure 3. Expression of neuronal markers in Ascl1 and Neurog2 iNs (corrected)

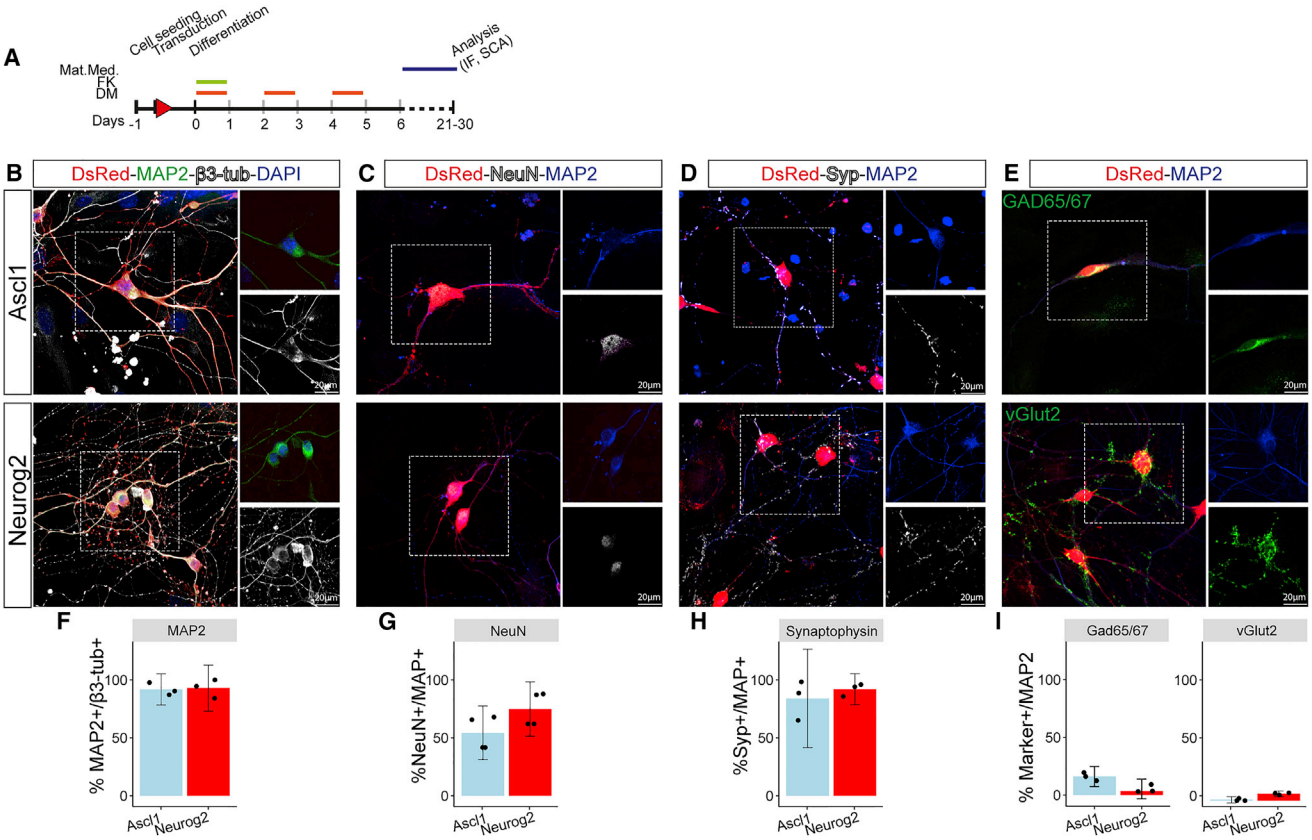


Figure 3. Expression of neuronal markers in Ascl1 and Neurog2 iNs (original)

DISCUSSION

In the present work, we aimed to shed light on several fundamental concepts that underlie the process of cell fate transition. In the first part, we uncovered how cell fate is erased by different reprogramming factors and what role reprogramming factor expression levels play in this. Furthermore, we leveraged the power of machine learning to establish whether conversion is a deterministic process directly linked to reprogramming factor expression levels. Finally, we also revealed a possible mechanism via which different reprogramming factors functionally interact with one another to resolve cell identity conflicts. In the second part, we expanded on this work by extending our scope to a different starter cell. In doing so, we were able to determine to what extent the cellular environment shapes the initial actions of reprogramming factors. Furthermore, we demonstrated that the final reprogrammed cellular identity is also heavily influenced by the developmental origin of the starter cell and the function of the reprogramming factor in this context. We believe that his knowledge will be useful to aid future endeavors aimed at improving current reprogramming strategies and facilitate its transition into the field of regenerative medicine.

General principles of fate conversion

To identify some of the general principles of reprogramming, we aimed to compare the actions of different reprogramming factors as well as analyze the result of their collision. To this end, we first developed a novel scRNA-seq paradigm, which we named Collide-seq. In contrast to most viral based strategies, Collide-seq offers detectable and quantifiable transgene expression due to the use of PiggyBacs (Mortazavi *et al.*, 2008). In addition, another considerable advantage of Collide-seq is that the use of inducible transgene expression and distinct fluorescent reporters allows the pooling of different conditions and the simultaneous initiation of fate conversion under identical conditions. As a result, we were able to systematically compare the actions of different reprogramming factors, something that to date has only been done to optimize the reprogramming into a defined target cell (Protze *et al.*, 2012; Yang *et al.*, 2019). This revealed that the induction of reprogramming factor expression rapidly resulted in switching off the original fibroblast identity for of all the factors we tested (Ascl1, MyoD1, FoxA2, Sox2), except Oct4. Interestingly, even though all factors downregulated genes related to fibroblast function, we found very little overlap between all four factors. In support of this, we found a similar result in our astrocyte experiments, where both of the factors we tested (Ascl1 & Neurog2) also shut down the original astrocyte identity via mostly different pathways. This seems to suggest that, although the removal of starter cell identity is a common feature

amongst all reprogramming factors, the mechanisms and dynamics are specific to a given reprogramming factor. However, from our first Collide-seq experiment, performed at 72h after transgene induction, we could not fully exclude the possibility that these factors may still use similar mechanisms and only differ in the temporal pattern they utilize these. This hypothesis is supported by the fact that we found substantial pairwise overlap between downregulated genes in this Collide-seq experiment. Data from our second Collide-seq experiment argues against this, however, as the inclusion of 24 and 48h time points did not substantially increase the amount of overlap. Thus, even with increased temporal resolution, we found a substantial fraction of downregulated genes that were unique to the factors tested, strongly arguing against the possibility that factors use similar mechanisms at different time points. This strongly suggests that downregulation of starter cell identity is a common characteristic of different reprogramming factors, though not achieved through a common mechanism.

Despite the shared capacity to downregulate starter cell identity between factors, we did find strong differences in potency between the different factors. While most factors showed a similar capacity to remove the existing identity and impose their own, the mesodermal factor MyoD1 displayed an exceptional potency in both aspects. We deemed this particularly interesting as the starting populations of MEFs was also of a mesodermal origin. Hence, the exceptional potency in this starter cell might be related to the developmental kinship between the two entities. This would be in line with previous data, showing that MyoD1 is inefficient in inducing the myogenic identity upon cells of a non-mesodermal origin (Davis, Weintraub and Lassar, 1987; Weintraub *et al.*, 1989). Along the same lines, the neurogenic reprogramming factor Neurog2 can also only efficiently convert ectoderm derived cortical astrocytes into neurons while it needs additional factors or small molecules to achieve neuronal conversion in mesoderm derived fibroblast (Berninger *et al.*, 2007; Heinrich *et al.*, 2010; Smith *et al.*, 2016). Taken together, these results support the hypothesis that direct conversion within a single germ layer would be more efficient and thorough than reprogramming between germ layers (Hochedlinger and Plath, 2009; Morris and Daley, 2013). This is surprising, as all factors tested here have been described as pioneer transcription factors. As such, differences in the epigenomic landscape, i.e., the accessibility of target sites, should be of little impact due to the ability of these factors to interact with closed chromatin. A possible explanation for these results, therefore, is that additional lineage conversion factors needed to drive conversion through cooperative chromatin binding (**Figure 7C**) are less abundant in distantly related cell types. This would make sense from a biological perspective, as factors driving alternative lineage gene expression are likely silenced in distantly related cell types due to their possibly detrimental effects on the maintenance of identity. The opposite might be the case in more closely related cell types, such

as astrocytes and neurons, where gene expression patterns have been shown to not be fully mutually exclusive and possibly facilitate conversion (Cahoy *et al.*, 2008; Zhang *et al.*, 2014; Herrero-Navarro *et al.*, 2021). Overall, the evidence presented here suggests that restricting direct conversion to cell types of the same germ layer might be preferable if remnants of the original identity are undesired, such as in clinical settings.

Finally, due to the use of PiggyBac constructs that could be detected with scRNA-seq, we were able to add another level of complexity to our dataset by determining the influence of expression levels on fate changes. This demonstrated that the expression and repression of target genes depends on a critical level of reprogramming factor expression. That is, target genes show a binary expression pattern in relation to reprogramming factor levels. Below the threshold, genes of the new fate are not induced and starter cell genes remain expressed. Above the threshold, the genes corresponding to the new program are induced while starter cell genes get switched off. Hence, expression levels either below or above the threshold are of little relevance as they do not substantially influence gene expression changes. This also suggested that there may be determinism of cellular states driven by gene expression levels. In other words, the expression levels of reprogramming factors might predict the array of different cell states as well as their stability. Indeed, we found there is global determinism in Collide-seq as the gene regulatory networks causal to cell fate could be learned with supervised machine learning models. Hence, this provides a proof of principle for the use of model-aided experimental design to achieve a desired outcome in reprogramming, provided that enough data is available on the actions of individual reprogramming factors.

Resolving cell identity conflicts

From the above, one might expect dominance of programs that are developmentally most close to the starter cell over more distantly related programs when instructed upon a cell simultaneously. This is, however, not what we observed. Instead, we found that the pairwise collision of reprogramming factors mostly results in the emergence of novel transcriptomic states that are not just the sum of both programs and which we refer to as collision states. Thus, although the instruction of a single fate seems to be most efficient when reprogramming into a developmentally related cell type, collision rather seems to be played out on the factor level. This is the first time that this question has been addressed in such a systematic matter. Conceptually, cell fusion experiments are perhaps the nearest equivalent (Cowan *et al.*, 2005; Brown and Fisher, 2021). As described above, cell fusion entails the fusion of a pluripotent cell to a differentiated somatic cell leading to the establishment of heterokaryons, a cell containing two distinct nuclei. As these experiments often induced hallmarks of pluripotency in the

resulting cells (Pralong, Trounson and Verma, 2006; Malinowski and Fisher, 2016; Imai *et al.*, 2020), one could imagine pluripotency to be dominant over terminally differentiated cell states. However, we did not find such dominance in our data, although the pluripotency factors (Oct4 and Sox2) in our panel have been reported to be sufficient to instruct this program when expressed together (Huangfu *et al.*, 2008; Kim *et al.*, 2008; Nemajerova *et al.*, 2012). Alternatively, it has been suggested that the brief induction of pluripotency is beneficial to the efficiency of direct reprogramming (Deleidi *et al.*, 2011; Kim *et al.*, 2011; Peskova *et al.*, 2019; Sharma *et al.*, 2019). Again, however, our data showed no indications that pluripotency factors act as general enablers of cell identity changes. The explanation for our inability to replicate these findings is most likely twofold. Under conventional culture conditions, when reprogramming with all four pluripotency factors, a fully stable pluripotent state is expected after 8 to 10 days (Stadtfeld *et al.*, 2008). Furthermore, most studies utilizing a transient pluripotency approach first observe a partially reprogrammed state after 4 days when using all factors (Margariti *et al.*, 2012; Karamariti *et al.*, 2013) and after 6 days when using only Oct4 (Peskova *et al.*, 2019). Our latest time point of 72h is therefore presumably too early to induce a pluripotent state, particularly with only one or two of the original four pluripotency factors. Additionally, our current culture conditions are likely unfavorable to the induction of pluripotency, as many studies use defined culture conditions containing several cytokines and small molecules to promote reprogramming (Efe *et al.*, 2011; Kim *et al.*, 2011; Mitchell *et al.*, 2014; Maza *et al.*, 2015). Taken together, a later time point of analysis or adaptation of culture conditions to facilitate the pluripotency factors might be necessary to fully address this matter. The former option, however, would be impractical in our setting as the reprogramming of MEFs to, for example, myoblasts, is rapid and shows drastic morphological and gene expression changes already after 3 days. Hence, a repetition of the current experiment under different culture conditions might be more feasible to fully address whether pluripotency factors can act as general enablers of cell fate.

To relate the outcome of factor competition to the relative reprogramming factor expression levels, we devised a novel analysis tool that we referred to as fate titration analysis (FTA). Briefly, FTA considered every cell as its own experiment and related the relative expression level of the two competing factors to a cell state outcome on the cluster level. This analysis showed that, when two factors collide and form a collision state (see above), competition is relatively independent of expression levels. That is, higher expression levels of one factor do not necessarily lead to the cell adopting that particular fate during competition. Instead, we found that the collision state is formed over virtually the entire spectrum of expression. This fitted well to critical threshold we determined for the expression and repression of target genes:

When reprogramming factors are expressed, they seem to only need to attain a certain threshold of expression to induce their program with little additive effect of exceeding this threshold. When they are competing, this results in the formation of a collision state that is insensitive to relative expression levels for the aforementioned reason. Instead, other mechanisms are likely at play that influence the outcome of this fate collision, which we aimed to unravel for one our collision pairs (see below). However, it has to be noted that regardless of the mechanism that gives rise to collision states, which may very well be different for different factor combinations, we generally found these states to be viable. That is, we found no indication that any of the multi-factor combinations suffered from increased cellular stress, as measured by the fraction of mitochondrial reads per condition in our single cell data. Furthermore, we find all multi-factor combinations to be present in our different Collide-seq experiments and found no clear loss of multi-factor conditions over time. More importantly, we found that at both 3- and 7-days post induction (See manuscript for day 3 data and **Figure 9** for day 7 data), cells double positive for the neurogenic factor Ascl1 and the myogenic factor MyoD1 can be found to survive *in vitro*. This strongly indicates that reprogramming factor collision results in semi-stable collision states that are viable at least one week after the induction of colliding programs.

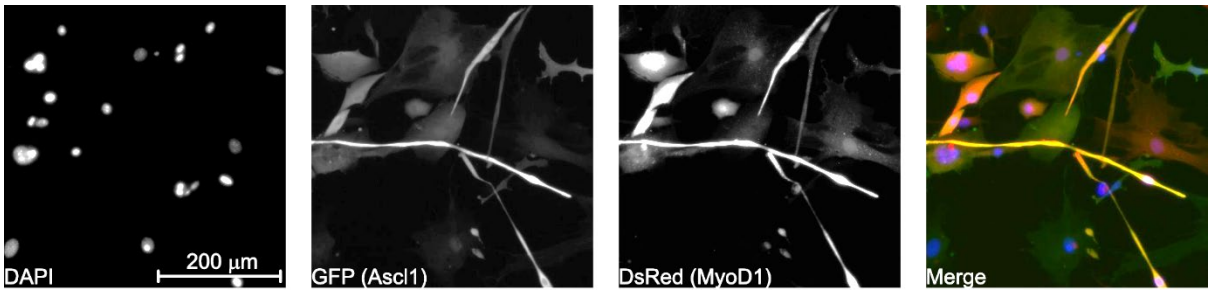


Figure 9: Stability of collision states. Shown are representative micrographs of collision between Ascl1 (GFP) and MyoD1 (DsRed) 7 days after induction of transgene expression. Scale bar represents 200 µm.

Mechanisms of fate collision

As a final part of our Collide-seq experiments, we aimed to shed light on the mechanisms behind fate collision. Although we deemed all factor collisions potentially interesting and worth investigating, we here decided to specifically follow up on the collision between the myogenic and neurogenic programs, i.e., Ascl1 and MyoD1. Mainly, this was due to the remarkable potency that Ascl1 displayed in perturbing MyoD1, the most potent individual transcription factor in our tested panel. In short, we found that Ascl1 is transcriptomically dominant over MyoD1, which was also reflected *in vitro* by the decreased formation of multinucleated

myoblasts upon collision. Employing a DNA binding mutant of Ascl1 led to a similar phenotype, suggesting that DNA binding and the subsequent collision of transcriptomic programs was unlikely to be the underlying mechanism of this competition (**Figure 10**). Instead, we found that both forms of Ascl1 reduced MyoD1 pioneer activity by preventing DNA binding and subsequent chromatin opening as well as gene expression. Overall, this deemed a more direct competition between both factors on the protein level to be the most likely cause of the observed effects (**Figure 10**). In this regard, it is important to consider that both factors belong to the family of bHLH transcription factors. bHLH factors generally function as either homo- or heterodimers (Massari and Murre, 2000; Murre, 2019), dimerizing either with E-proteins, Id proteins or other bHLH factors to bind to canonical E-box motifs and drive transcription (Massari and Murre, 2000; Henke *et al.*, 2009; Murre, 2019). Given that the use of mutAscl1 made transcriptional and DNA binding site competition unlikely, we hypothesized that both factors might either form non-functional heterodimers that bind to the DNA but do not activate transcription or one factor would sequester E-proteins from the other (**Figure 10**). However, as mentioned above, we observed a clear loss of MyoD1 DNA binding upon competition, making the formation of non-functional DNA binding heterodimers unlikely. Therefore, the competition for cofactors seemed to be the most plausible explanation. Evidence to support this notion came from known motif analysis on MyoD1 binding sites. This revealed that, upon competition, there is a substantial reduction of the fraction of binding sites that have an E47/Tcf3 motif. This suggests that the unavailability of the E-protein Tcf3 to MyoD1 might impair its DNA binding and thereby cause the loss of pioneer factor activity. As such, this provides a tentative mechanism for the competition between Ascl1 and MyoD1, similar to what has been reported before for MyoD1 and Twist (Spicer *et al.*, 1996). In the end, however, it is important to consider that the above-described mechanisms might only apply to these two factors, while additional mechanisms might exist for other factors. Indeed, this seems likely as we found collision states emerging for many of our factor pairs though they belonged to different families of transcription factors. As such, the results described here should be considered as a first indication of how cells resolve cell identity conflicts and cofactor competition is most likely only one of those ways.

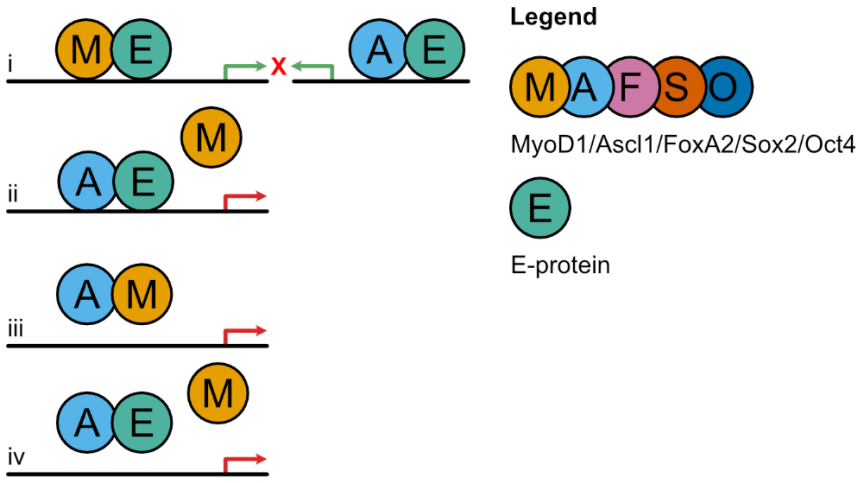


Figure 10: Possible Ascl1-MyoD1 collision mechanisms. i) Collision of transcriptional programs, ii) DNA binding site competition, iii) Formation of non-functional heterodimers and iv) competition for E-proteins.

Influence of starter cell on reprogramming

In addition to comparing different factors within the same cellular environment, we also aimed to determine how the identity of the starter cell influences the early stages of reprogramming as well as its outcome. To this end, we first optimized the isolation of astrocytes, as we required a highly pure population to address this question. We purified cells based on the expression off known astrocyte markers. Although it would have been possible to use fluorescent reporter mice and fluorescence-activated cell sorting to achieve this, this method has been known to produce large amounts of sheer stress that might impede further culture (Guttenplan and Liddelow, 2019). Hence, we employed a magnetic-activated cell sorting strategy based on the expression of anti-astrocyte cell surface antigen-2 (Acsa-2) that involved a milder dissociation and sorting procedure (G. Kantzer *et al.*, 2017). As a proof of principle, we also confirmed that the cells obtained with this procedure where predominantly astrocytes (**Figure 11**).

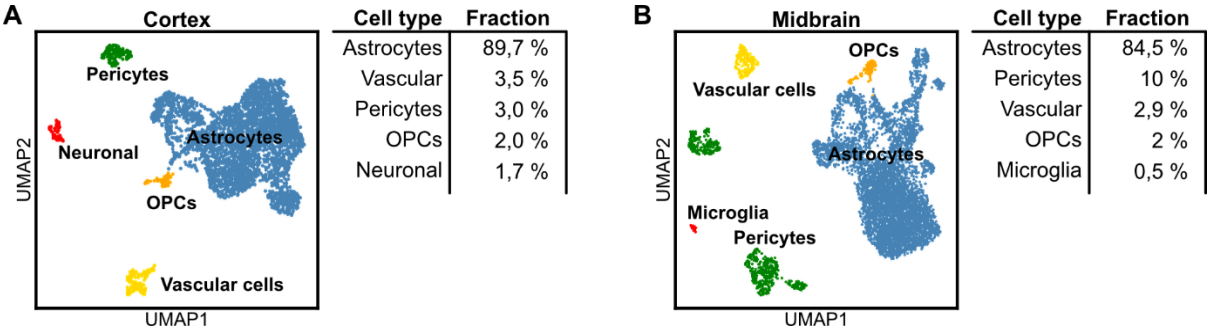


Figure 11: Verification of astrocyte purity after magnetic-activated cell sorting. Shown are 2D embeddings using Uniform Manifold Approximation and Projection (UMAP) of single cell RNA sequencing data of sorted cortical (A) and midbrain (B) astrocytes using anti-astrocyte cell surface antigen-2. VECs: Vascular endothelial cells, OPCs: Oligodendrocyte progenitor cells.

The next step was to reprogram these regional astrocytes into neurons with Ascl1 and Neurog2. In doing so, we noted that both factors instruct very different programs in a starting population of spinal cord astrocytes. Furthermore, both factors also instructed very different programs as compared to the ones they instruct in cortical astrocytes (Masserdotti *et al.*, 2015). In cortical astrocytes, Ascl1 instructs a GABAergic program while Neurog2 yields neurons of a glutamatergic subtype (Heinrich *et al.*, 2010). In spinal cord astrocytes, however, we did not find signs of a specification towards these subtypes, in contrast to recent reports (Hu *et al.*, 2019). Instead, their neurogenic cascades converged on the generation of GABAergic ventral spinal cord V2b interneurons and their p2 progenitors. Remarkably, this finding fits well to the role these factors play during neural tube development, as both factors are expressed in p2 progenitors in a salt and pepper like manner (Misra *et al.*, 2014). Therefore, these results provided a clear indication that the starter cell heavily influences the behavior of potent neurogenic reprogramming factors and is in line with the current literature (Chouchane *et al.*, 2017; Herrero-Navarro *et al.*, 2021; Rao *et al.*, 2021). In support of this notion, we also found that key patterning genes from the Hox family of transcription factors, which specify spinal cord domains during development, were expressed in both acutely isolated and cultured astrocytes (Carpenter, 2002; Philippidou and Dasen, 2013). More importantly, they continued to be expressed in induced neurons obtained from spinal cord astrocytes, supporting their role in defining cellular identity and fitting to the observation that patterning factor Gbx2 was sufficient to redirect reprogramming to a different regional neuron identity (Herrero-Navarro *et al.*, 2021). Taken together, we have provided strong evidence that the (regional) identity of the starter cell shapes the actions of (neuronal) reprogramming factors, in line with the correct layer specificity of *in vivo* reprogrammed astrocytes (Mattugini *et al.*, 2019).

CONCLUDING REMARKS

One of the major open questions in cell biology is what defines a cellular identity. As described above, although the advent of techniques to profile the transcriptome at the single cell level has helped cataloguing diversity, how identity is regulated remains relatively unknown. This is particularly important in the context of reprogramming, as the final aim is to generate fully functional cell types that closely resemble their endogenous counterparts. That is, cells that do not fulfill one specific function, but instead reproduce the entire cellular phenotype and show the capacity to respond and adapt to changes in their environment. Only when this is achieved can we think about achieving functional recovery of damaged tissues through the replacement of lost cells with reprogramming strategies. Towards achieving this goal, the current work provided several key findings to facilitate this process.

For example, it is of crucial importance for reprogramming to understand how the original starter cell identity is downregulated, as remnants of starter cell identity may have detrimental effects on cell function. In the present work we demonstrate that, although the loss of starter cell identity is commonly achieved by different reprogramming factors, it most efficiently occurs when the conversion occurs between cell types that belong to the same germ layer. Hence, to avoid the maintenance of starting cell identity, developmentally closer related conversions are preferable. In addition, specific factors that aid the downregulation of starter cell identity may also be discovered in future studies. Although we looked for such factors in our current dataset, in particular within our most potent factor condition, we were unable to identify these. As the current single cell methods still capture only a fraction of the entire transcriptome (Haque *et al.*, 2017), further technological improvements are likely required to identify such key players.

Another important step is to continue to optimize current reprogramming protocols as to maximize their efficiency and reduce any negative side effects such as cell stress or even cell death. To this end, optimizing reprogramming factor levels upon ectopic expression may prove fruitful, in addition to the already demonstrated benefits of reducing metabolic stress (Russo *et al.*, 2021). With regard to this, our work revealed that a reprogramming factor needs to reach a certain expression level threshold to initiate fate conversion. It is likely that most currently available reprogramming paradigms, including Collide-seq, exceed this threshold due to the use of strong viral promoters that drive gene expression. Hence, titrating the reprogramming factor levels so that they only just satisfy the threshold might improve reprogramming efficiency and the general condition of reprogrammed cells due to reduced stress on the transcriptional machinery and the cell in general.

The application of artificial intelligence, and in particular machine-learning, on the ever-growing amount of (single cell) datasets provides an alternative way through which reprogramming paradigms can be designed and optimized. Indeed, reports of the application of such techniques have recently started to be published (Kamimoto, Hoffmann and Morris, 2020). Here, we describe determinism of cell states with regard to reprogramming factor expression levels to provide a further proof of principle that these rules can be abstracted and learned with machine learning. In turn, this may aid the development of model-based experimental design strategies aimed at the better-informed selection of factors necessary to attain a certain cell state. This might not only improve efficiency, but also the overall fidelity of the generated cell types, thereby providing a novel way of optimizing current reprogramming paradigms.

A final consideration we would like to stress with regard to achieving optimal reprogramming outcomes is that of starter cell selection. In particular, this poses an important consideration for the reprogramming of cells within the CNS, as it shows a higher degree of regionalization than other organs (Lake *et al.*, 2016; Saunders *et al.*, 2018; Herrero-Navarro *et al.*, 2021). Furthermore, there is a tremendous degree of neuronal diversity within the mammalian brain that is often intricately related to this regionalization (Zeisel *et al.*, 2015, 2018; Mancinelli and Lodato, 2018; Saunders *et al.*, 2018; Özel *et al.*, 2020). In this regard, astrocytes represent an ideal *in loco* population of cells to replace lost neurons. Furthermore, several lines of evidence have demonstrated that a regional identity is also conferred upon astrocytes and this regionalization facilitates the reprogramming into appropriate neuronal subtypes (Tsai *et al.*, 2012; Chai *et al.*, 2017; Chouchane *et al.*, 2017; John Lin *et al.*, 2017; Boisvert *et al.*, 2018; Lanjakornsiripan *et al.*, 2018; Hu *et al.*, 2019; Mattugini *et al.*, 2019; Herrero-Navarro *et al.*, 2021; Rao *et al.*, 2021). Our work confirmed these findings and demonstrated that the same factors initiate very different programs depending on the regional origin of the starter cell. Furthermore, they can lead to the generation of a regionally specified neuronal subtype. Although this is encouraging when aiming to replace lost cells, it might pose a problem for strategies aimed at generating specific subtypes of neurons in abnormal locations. For example, generating dopaminergic neurons in the striatum to counteract the loss of dopamine afferents in this region after Parkinson's disease.

In summary, by investigating the mechanisms of direct conversion, the results of my Ph.D. thesis warrant a more holistic view on the reprogramming process, which is not merely the induction of new genes but requires crosstalk between starter cell identity and the reprogramming factor to dramatically modify the intracellular properties and morphology of the newly established identity.

ACKNOWLEDGEMENTS

Before moving to the last few formally required pages, I would like to thank several people. First, my primary supervisor Magdalena. Thank you for your mentoring and guidance throughout the past years. Thank you as well for the exceptional scientific possibilities you provide us with in the lab, which have elevated my projects. I would also like to thank you for your understanding during the more difficult times, giving me time and space when I needed it. Finally, I would like to express my appreciation for your unparalleled enthusiasm for science and my gratefulness for having the possibility to learn from you during my PhD.

Similarly, I would like to thank my co-supervisor Giacomo. I honestly cannot thank you enough for all the input and help you have given me. Whether it was naive questions on things I should have known, reading my thesis numerous times or discussing about possible experiments, your door was always open and I am truly grateful for that. I look back at our working relationship with fondness and will cherish the memory of our trip to Rome for a long time. Along the same lines, I would also like to thank Stefan for co-supervising my work and always thinking outside the box and coming up with original ideas and hypotheses. Finally, I would like to extend my thanks to all lab members and technicians who have in any way helped me over the past few years. In particular, I am sincerely thankful for all the help and patience Tatiana has given me, becoming partners in crime during countless 10x experiments. All of you have been an instrumental part of my PhD, thank you!

I would like to continue by thanking a few people on a more personal level. First of all, I would like to thank Matteo “Lo Stronzo” P. and Poornemaa “Ninja” Natarajan, who have not only been great colleagues but have also become close friends. I know the former will deny this fact, but I’ve come to terms with his denial. Truly, your friendship means a lot to me, and I am forever grateful to have met and worked with both of you. Matteo, you’re a glorious mess in the best imaginable way. You’re an extremely talented scientist and a master of drama, astrology and general sassiness. Stay as you are and I hope we meet one day on Sicily while I’m trying to drive a Vespa. Poornemaa, I’m so happy we managed to break your shell and have you in my life. You’re wiser and way more talented than you give yourself credit for and don’t let anyone else tell you otherwise. I look forward to renting a Vespa with you and proving Matteo wrong, even if it’s only until the end of the street.

Staying within the lab, I would also like to thank Giulia A., Florencia M. and Giorgia B. for the great fun we had during dinners and drinks. Similarly, I would like to thank Sonia N. for her friendship and support during the first years of my PhD. Although, you left me/us for love, I highly appreciate the time we spent together in Munich and beyond. Shifting labs a little, I

would also like to thank the members of the Stricker lab. In particular, I would like to thank Christopher “Big Boy” B., otherwise known as the backpfeifengesicht. You were the first person to welcome me in the lab and we had a bunch of cool adventures afterwards that I’m super grateful for. I would also like to thank Maxi W. for, well, food mainly. Not only grabbing me lunch but even more so the countless recommendations for restaurants and other food places in Munich, particularly Turkitch and Mai Garten. However, let’s also not forget the many video game and pizza discussions that elevated my palate. Thanks Maxi! Finally, I would also like to thank Anna K. for shared ranting about Munich, regular Turkitch trips and commuting together. I hope to visit you soon in Vienna! Changing floors one more time, I would also like to thank Klara “Lady KTN” N. for the many hiking trips and dinners we shared!

Extending beyond confines of the lab, I would like to thank my flatmates. Ihr seid alle behindert, aber ich liebe euch trotzdem. Stefan, du bist eine ❓🟫🚑✨✔️✔️❗🚾🎁 vergiss das bitte nicht. Sexy, es war eine Ehre, unter deiner Lordschaft zu leben. Linze, vergiss bitte nicht, vorsichtig zu sein. Das kann gefährlich sein. Klarischmari, ich drücke die Daumen, dass Kimmich eines Tages Single wird.

Next, I would like to thank a few people from back home. In particular, Hannah G., Kayleigh T., Boyd W., Robbelien K., Sterre O., and Margherita F. Hannah bedankt voor de ontelbare uren die we met elkaar aan de telefoon gehangen hebben. Ik weet nog goed dat je goede voornemen voor, waarschijnlijk was het 2017, vaker bellen was. Ik ben heel blij dat we dat inmiddels zo’n vier jaar hebben volgehouden en dat we elkaar nog steeds regelmatig spreken. Onze vriendschap is me enorm dierbaar en daarvoor bij deze dank. Kayleigh, voor jou gaat bijna hetzelfde op. Ik denk dat we van geluk mogen spreken dat we tegenwoordig geen telefoonrekeningen meer krijgen, anders waren we flink de pineut geweest. Bedankt voor je steun en luisterend oor door dik en dun. Boyd, je bent een man van weinig woorden maar bedankt voor de game sessies en afleiding die dit met zich meebracht. Robba, ook jou heb ik veel aan de telefoon gehad, met name toen we allebei net aan onze PhD begonnen waren. Het is voor mij altijd heel waardevol geweest met iemand in een vergelijkbare situatie te kunnen sparren, bedankt daarvoor. Sterre, we hebben misschien niet zoveel gebeld, maar ik heb je de afgelopen paar jaar denk ik het vaakst in München mogen ontvangen. Helaas heeft een niet nader te noemen ziekteverwekker de laatste twee jaar wat roet in het eten gegooid, maar ik hoop je op een volgende bestemming weer even vaak te mogen verwelkomen. Okay, fair enough Marghe, you’re not Dutch and don’t even live in the Netherlands anymore but my memories of you will always be connected to the Netherlands. Anyway, I’m very happy that we managed to go from the status of an unwanted flyer that gets handed to you while shopping to a friendship that has lasted for several years. Even though getting a hold of each other hasn’t always been easy, I greatly appreciate that our friendship has persisted after all these years.

Last but not least, I would like to thank my parents, brother and grandparents for their support throughout the years. I wouldn't have achieved all that I have without your unconditional love. Thank you for always being only one phone call away when I needed it and coming to visit despite the endless number of baustellen you have to pass to get here. I would also like to thank my newfound Polish family. Dziękuję za tak miłe przyjęcie. Finally, I would like to thank Aleksandra "Ola" T. for her love and support. You have made the final stretch so much easier.

REFERENCES

The references below serve as a bibliography to the introduction and discussion of the thesis. The references used in the results section are provided in their respective sections.

Allen, N.J. and Barres, B.A. (2009) "Glia — more than just brain glue," *Nature 2009 457:7230*, 457(7230), pp. 675–677. doi:10.1038/457675a.

Allen, N.J. and Eroglu, C. (2017) "Cell Biology of Astrocyte-Synapse Interactions," *Neuron*, 96(3), pp. 697–708. doi:10.1016/J.NEURON.2017.09.056.

Allen, N.J. and Lyons, D.A. (2018) "Glia as architects of central nervous system formation and function.," *Science (New York, N.Y.)*, 362(6411), pp. 181–185. doi:10.1126/science.aat0473.

Alvarez-Dolado, M. *et al.* (2003) "Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes," *Nature 2003 425:6961*, 425(6961), pp. 968–973. doi:10.1038/nature02069.

Anthony, T.E. *et al.* (2004) "Radial Glia Serve as Neuronal Progenitors in All Regions of the Central Nervous System," *Neuron*, 41(6), pp. 881–890. doi:10.1016/S0896-6273(04)00140-0.

Araque, A. *et al.* (1999) "Tripartite synapses: glia, the unacknowledged partner," *Trends in neurosciences*, 22(5), pp. 208–215. doi:10.1016/S0166-2236(98)01349-6.

Azam, S. *et al.* (2021) "The Ageing Brain: Molecular and Cellular Basis of Neurodegeneration," *Frontiers in Cell and Developmental Biology*, 9, p. 2120. doi:10.3389/fcell.2021.683459.

Babeu, J.P. and Boudreau, F. (2014) "Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks," *World Journal of Gastroenterology*, 20(1), pp. 22–30. doi:10.3748/wjg.v20.i1.22.

Bagley, J.A. *et al.* (2017) "Fused cerebral organoids model interactions between brain regions," *Nature Methods 2017 14:7*, 14(7), pp. 743–751. doi:10.1038/nmeth.4304.

Baker, N.E. and Brown, N.L. (2018) "All in the family: proneural bHLH genes and neuronal diversity," *Development (Cambridge, England)*, 145(9). doi:10.1242/DEV.159426.

Bandler, R.C. *et al.* (2021) "Single-cell delineation of lineage and genetic identity in the mouse brain," *Nature 2021 601:7893*, 601(7893), pp. 404–409. doi:10.1038/s41586-021-04237-0.

Barker, R.A. *et al.* (2017) "Human Trials of Stem Cell-Derived Dopamine Neurons for Parkinson's Disease: Dawn of a New Era," *Cell Stem Cell*, 21(5), pp. 569–573. doi:10.1016/J.STEM.2017.09.014.

Barlow, P., Owen, D.A. and Graham, C. (1972) "DNA synthesis in the preimplantation mouse embryo," *Development*, 27(2), pp. 431–445. doi:10.1242/DEV.27.2.431.

Barres, B.A. (2008) "The Mystery and Magic of Glia: A Perspective on Their Roles in Health and Disease," *Neuron*, 60(3), pp. 430–440. doi:10.1016/J.NEURON.2008.10.013.

Batiuk, M.Y. *et al.* (2020) "Identification of region-specific astrocyte subtypes at single cell resolution," *Nature Communications 2020 11:1*, 11(1), pp. 1–15. doi:10.1038/s41467-019-14198-8.

Bayraktar, O.A. *et al.* (2020) "Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map," *Nature Neuroscience 2020 23:4*, 23(4), pp. 500–509. doi:10.1038/s41593-020-0602-1.

Ben-Arie, N. *et al.* (1996) "Evolutionary Conservation of Sequence and Expression of the bHLH Protein Atonal Suggests a Conserved Role in Neurogenesis," *Human Molecular Genetics*, 5(9), pp. 1207–1216. doi:10.1093/HMG/5.9.1207.

Berninger, B. *et al.* (2007) "Functional Properties of Neurons Derived from In Vitro Reprogrammed Postnatal Astroglia," *Journal of Neuroscience*, 27(32), pp. 8654–8664. doi:10.1523/JNEUROSCI.1615-07.2007.

Bertrand, N., Castro, D.S. and Guillemot, F. (2002) "Proneural genes and the specification of neural cell types," *Nature Reviews Neuroscience*, 3(7), pp. 517–530. doi:10.1038/nrn874.

Bier, E. and de Robertis, E.M. (2015) "EMBRYO DEVELOPMENT. BMP gradients: A paradigm for morphogen-mediated developmental patterning.," *Science (New York, N.Y.)*, 348(6242), p. aaa5838. doi:10.1126/science.aaa5838.

Birey, F. *et al.* (2017) "Assembly of functionally integrated human forebrain spheroids," *Nature 2017 545:7652*, 545(7652), pp. 54–59. doi:10.1038/nature22330.

Bishop, K.M., Goudreau, G. and O'Leary, D.D. (2000) "Regulation of area identity in the mammalian neocortex by Emx2 and Pax6.," *Science (New York, N.Y.)*, 288(5464), pp. 344–9. doi:10.1126/science.288.5464.344.

Blau, H.M. and Blakely, B.T. (1999) "Plasticity of cell fate: Insights from heterokaryons," *Seminars in Cell & Developmental Biology*, 10(3), pp. 267–272. doi:10.1006/SCDB.1999.0311.

Boisvert, M.M. *et al.* (2018) "The Aging Astrocyte Transcriptome from Multiple Regions of the Mouse Brain," *Cell Reports*, 22(1), pp. 269–285. doi:10.1016/j.celrep.2017.12.039.

Brattås, P.L. *et al.* (2021) "Impact of differential and time-dependent autophagy activation on therapeutic efficacy in a model of Huntington disease," *Autophagy*, 17(6), pp. 1316–1329. doi:10.1080/15548627.2020.1760014.

Braun, T. *et al.* (1992) "Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death," *Cell*, 71(3), pp. 369–382. doi:10.1016/0092-8674(92)90507-9.

Braun, T. and Arnold, H.H. (1995) "Inactivation of Myf-6 and Myf-5 genes in mice leads to alterations in skeletal muscle development.," *The EMBO journal*, 14(6), pp. 1176–86. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7720708 (Accessed: November 24, 2021).

Briançon, N. *et al.* (2004) "Expression of the alpha7 isoform of hepatocyte nuclear factor (HNF) 4 is activated by HNF6/OC-2 and HNF1 and repressed by HNF4alpha1 in the liver.," *The Journal of biological chemistry*, 279(32), pp. 33398–408. doi:10.1074/jbc.M405312200.

Briggs, R. and King, T.J. (1952) "Transplantation of living nuclei from blastula cells into enucleated frogs' eggs," *Proceedings of the National Academy of Sciences*, 38(5), pp. 455–463. doi:10.1073/pnas.38.5.455.

Brown, K.E. and Fisher, A.G. (2021) "Reprogramming lineage identity through cell–cell fusion," *Current Opinion in Genetics & Development*, 70, pp. 15–23. doi:10.1016/J.GDE.2021.04.004.

Buganim, Y. *et al.* (2012) "Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase.," *Cell*, 150(6), pp. 1209–22. doi:10.1016/j.cell.2012.08.023.

Bush, T.G. *et al.* (1999) "Leukocyte infiltration, neuronal degeneration, and neurite outgrowth after ablation of scar-forming, reactive astrocytes in adult transgenic mice.," *Neuron*, 23(2), pp. 297–308. doi:10.1016/s0896-6273(00)80781-3.

Bushong, E.A. *et al.* (2002) "Protoplasmic astrocytes in CA1 stratum radiatum occupy separate anatomical domains," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 22(1), pp. 183–192. doi:10.1523/JNEUROSCI.22-01-00183.2002.

Cacchiarelli, D. *et al.* (2018) "Aligning Single-Cell Developmental and Reprogramming Trajectories Identifies Molecular Determinants of Myogenic Reprogramming Outcome.," *Cell systems*, 7(3), pp. 258-268.e3. doi:10.1016/j.cels.2018.07.006.

Cahoy, J.D. *et al.* (2008) "A Transcriptome Database for Astrocytes, Neurons, and Oligodendrocytes: A New Resource for Understanding Brain Development and Function," *Journal of Neuroscience*, 28(1), pp. 264–278. doi:10.1523/JNEUROSCI.4178-07.2008.

Cakir, B. *et al.* (2019) "Engineering of human brain organoids with a functional vascular-like system," *Nature Methods 2019 16:11*, 16(11), pp. 1169–1175. doi:10.1038/s41592-019-0586-5.

Callaway, E.M. *et al.* (2021) "A multimodal cell census and atlas of the mammalian primary motor cortex," *Nature*, 598(7879), pp. 86–102. doi:10.1038/s41586-021-03950-0.

Campbell, K.H.S. *et al.* (1996) "Sheep cloned by nuclear transfer from a cultured cell line," *Nature 1996 380:6569*, 380(6569), pp. 64–66. doi:10.1038/380064a0.

Cao, J. *et al.* (2019) "The single-cell transcriptional landscape of mammalian organogenesis," *Nature*, 566(7745), pp. 496–502. doi:10.1038/S41586-019-0969-X.

Carpenter, E.M. (2002) "Hox genes and spinal cord development," *Developmental neuroscience*, 24(1), pp. 24–34. doi:10.1159/000064943.

Casarosa, S., Fode, C. and Guillemot, F. (1999) "Mash1 regulates neurogenesis in the ventral telencephalon," *Development*, 126(3), pp. 525–534. doi:10.1242/DEV.126.3.525.

Cau, E., Casarosa, S. and Guillemot, F. (2002) "Mash1 and Ngn1 control distinct steps of determination and differentiation in the olfactory sensory neuron lineage," *Development*, 129(8), pp. 1871–1880. doi:10.1242/DEV.129.8.1871.

Cereghini, S. *et al.* (1992) "Expression patterns of vHNF1 and HNF1 homeoproteins in early postimplantation embryos suggest distinct and sequential developmental roles," *Development*, 116(3), pp. 783–797. doi:10.1242/DEV.116.3.783.

Chai, H. *et al.* (2017) "Neural Circuit-Specialized Astrocytes: Transcriptomic, Proteomic, Morphological, and Functional Evidence," *Neuron*, 95(3), pp. 531-549.e9. doi:10.1016/j.neuron.2017.06.029.

Chakraborty, S. *et al.* (2014) "A CRISPR/Cas9-Based System for Reprogramming Cell Lineage Specification," *Stem Cell Reports*, 3(6), pp. 940–947. doi:10.1016/J.STEMCR.2014.09.013.

Chanda, S. *et al.* (2014) "Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1," *Stem Cell Reports*, 3(2), pp. 282–296. doi:10.1016/J.STEMCR.2014.05.020.

Chazaud, C. *et al.* (2006) "Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway," *Developmental Cell*, 10(5), pp. 615–624. doi:10.1016/J.DEVCEL.2006.02.020.

Chen, J. *et al.* (2016) "Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming," *Cell Reports*, 14(6), pp. 1540–1554. doi:10.1016/j.celrep.2016.01.013.

Chouchane, M. *et al.* (2017) "Lineage Reprogramming of Astroglial Cells from Different Origins into Distinct Neuronal Subtypes," *Stem Cell Reports*, 9(1), pp. 162–176. doi:10.1016/j.stemcr.2017.05.009.

Chronis, C. *et al.* (2017) "Cooperative Binding of Transcription Factors Orchestrates Reprogramming.," *Cell*, 168(3), pp. 442-459.e20. doi:10.1016/j.cell.2016.12.016.

Clarke, B.E. *et al.* (2021) "Regionally encoded functional heterogeneity of astrocytes in health and disease: A perspective," *Glia*, 69(1), pp. 20–27. doi:10.1002/GLIA.23877.

Clarke, L.E. *et al.* (2018) "Normal aging induces A1-like astrocyte reactivity.," *Proceedings of the National Academy of Sciences of the United States of America*, 115(8), pp. E1896–E1905. doi:10.1073/pnas.1800165115.

Clotman, F. *et al.* (2002) "The onecut transcription factor HNF6 is required for normal development of the biliary tract," *Development*, 129(8), pp. 1819–1828. doi:10.1242/DEV.129.8.1819.

Cohen, M., Briscoe, J. and Blassberg, R. (2013) "Morphogen interpretation: the transcriptional logic of neural tube patterning," *Current Opinion in Genetics & Development*, 23(4), pp. 423–428. doi:10.1016/J.GDE.2013.04.003.

Condic, M.L. (2014) "Totipotency: What it is and what it is not," *Stem Cells Dev.*, 23(8), pp. 796–812. doi:10.1089/scd.2013.0364.

Constantinides, P.G., Jones, P.A. and Gevers, W. (1977) "Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment," *Nature 1977 267:5609*, 267(5609), pp. 364–366. doi:10.1038/267364a0.

Costelloe, E.O. *et al.* (1999) "Regulation of the plasminogen activator inhibitor-2 (PAI-2) gene in murine macrophages. Demonstration of a novel pattern of responsiveness to bacterial endotoxin," *Journal of Leukocyte Biology*, 66(1), pp. 172–182. doi:10.1002/JLB.66.1.172.

Cowan, C.A. *et al.* (2005) "Nuclear Reprogramming of Somatic Cells After Fusion with Human Embryonic Stem Cells," *Science*, 309(5739), pp. 1369–1373. doi:10.1126/SCIENCE.1116447.

Cubas, P. *et al.* (1991) "Proneural clusters of achaete-scute expression and the generation of sensory organs in the Drosophila imaginal wing disc," *Genes and Development*, 5(6), pp. 996–1008. doi:10.1101/GAD.5.6.996.

Cyranoski, D. (2018) "'Reprogrammed' stem cells implanted into patient with Parkinson's disease," *Nature* [Preprint]. doi:10.1038/d41586-018-07407-9.

Dall'Agnese, A. *et al.* (2019) "Transcription Factor-Directed Re-wiring of Chromatin Architecture for Somatic Cell Nuclear Reprogramming toward trans-Differentiation," *Molecular Cell*, 76(3), pp. 453-472.e8. doi:10.1016/j.molcel.2019.07.036.

Davies, P.S. *et al.* (2009) "Inflammation and proliferation act together to mediate intestinal cell fusion.," *PloS one*, 4(8), p. e6530. doi:10.1371/journal.pone.0006530.

Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) "Expression of a single transfected cDNA converts fibroblasts to myoblasts," *Cell*, 51(6), pp. 987–1000. doi:10.1016/0092-8674(87)90585-X.

DeBoever, C. (2017) "Large-scale profiling reveals the influence of genetic variation on gene expression in human induced pluripotent stem cells," *Cell Stem Cell*, 20, pp. 533–546.

Deleidi, M. *et al.* (2011) "Oct4-Induced Reprogramming Is Required for Adult Brain Neural Stem Cell Differentiation into Midbrain Dopaminergic Neurons," *PLoS ONE*. Edited by J.C. Zheng, 6(5), p. e19926. doi:10.1371/journal.pone.0019926.

Deneris, E.S. and Hobert, O. (2014) "Maintenance of postmitotic neuronal cell identity," *Nature neuroscience*, 17(7), pp. 899–907. doi:10.1038/NN.3731.

Deng, W. *et al.* (2021) "The transcription factor code in iPSC reprogramming," *Current Opinion in Genetics & Development*, 70, pp. 89–96. doi:10.1016/J.GDE.2021.06.003.

Dezonne, R.S. *et al.* (2013) "Thyroid hormone treated astrocytes induce maturation of cerebral cortical neurons through modulation of proteoglycan levels," *Frontiers in Cellular Neuroscience*, 7(JUL), p. 125. doi:10.3389/fncel.2013.00125.

Dimicoli-Salazar, S. *et al.* (2011) "Efficient in vitro myogenic reprogramming of human primary mesenchymal stem cells and endothelial cells by Myf5," *Biology of the cell*, 103(11), pp. 531–542. doi:10.1042/BC20100112.

Drögemüller, K. *et al.* (2008) "Astrocyte gp130 expression is critical for the control of Toxoplasma encephalitis," *Journal of immunology (Baltimore, Md. : 1950)*, 181(4), pp. 2683–2693. doi:10.4049/JIMMUNOL.181.4.2683.

Drouin-Ouellet, J. *et al.* (2017) "REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways," *EMBO Molecular Medicine*, 9(8), pp. 1117–1131. doi:10.15252/EMMM.201607471.

Du, Y. *et al.* (2014) "Human Hepatocytes with Drug Metabolic Function Induced from Fibroblasts by Lineage Reprogramming," *Cell Stem Cell*, 14(3), pp. 394–403. doi:10.1016/J.STEM.2014.01.008.

Dudek, R.W. (2009) *High-Yield Embryology*. 4th edn. Lippincot Williams & Wilkins.

Duelen, R. and Sampaolesi, M. (2017) "Stem Cell Technology in Cardiac Regeneration: A Pluripotent Stem Cell Promise," *EBioMedicine*, 16, pp. 30–40. doi:10.1016/J.EBIOM.2017.01.029.

Durkin, M.E. *et al.* (2013) "Isolation of Mouse Embryo Fibroblasts," *Bio-protocol*, 3(18). doi:10.21769/bioprotoc.908.

Dyce, J. *et al.* (1987) "Do trophectoderm and inner cell mass cells in the mouse blastocyst maintain discrete lineages?," *Development*, 100(4), pp. 685–698. doi:10.1242/DEV.100.4.685.

Eddleston, M. and Mucke, L. (1993) "Molecular profile of reactive astrocytes—Implications for their role in neurologic disease," *Neuroscience*, 54(1), pp. 15–36. doi:10.1016/0306-4522(93)90380-X.

Efe, J.A. *et al.* (2011) "Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy," *Nature Cell Biology 2011 13:3*, 13(3), pp. 215–222. doi:10.1038/ncb2164.

Eroglu, C. and Barres, B.A. (2010) "Regulation of synaptic connectivity by glia," *Nature 2010 468:7321*, 468(7321), pp. 223–231. doi:10.1038/nature09612.

Farah, M.H. *et al.* (2000) "Generation of neurons by transient expression of neural bHLH proteins in mammalian cells," *Development*, 127(4), pp. 693–702. doi:10.1242/DEV.127.4.693.

Faulkner, J.R. *et al.* (2004) "Reactive astrocytes protect tissue and preserve function after spinal cord injury," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(9), pp. 2143–2155. doi:10.1523/JNEUROSCI.3547-03.2004.

Faustino Martins, J.M. *et al.* (2020) "Self-Organizing 3D Human Trunk Neuromuscular Organoids," *Cell Stem Cell*, 26(2), pp. 172-186.e6. doi:10.1016/J.STEM.2019.12.007.

Femino, A.M. *et al.* (1998) "Visualization of single RNA transcripts in situ.," *Science (New York, N.Y.)*, 280(5363), pp. 585–90. doi:10.1126/science.280.5363.585.

Ferretti, E. and Hadjantonakis, A.K. (2019) "Mesoderm specification and diversification: from single cells to emergent tissues," *Current Opinion in Cell Biology*, 61, pp. 110–116. doi:10.1016/J.CEB.2019.07.012.

Figley, C.R. and Stroman, P.W. (2011) "The role(s) of astrocytes and astrocyte activity in neurometabolism, neurovascular coupling, and the production of functional neuroimaging signals," *European Journal of Neuroscience*, 33(4), pp. 577–588. doi:10.1111/J.1460-9568.2010.07584.X.

Fishman, V.S. *et al.* (2015) "Cell divisions are not essential for the direct conversion of fibroblasts into neuronal cells.," *Cell cycle (Georgetown, Tex.)*, 14(8), pp. 1188–96. doi:10.1080/15384101.2015.1012875.

Fleming, T.P. (1987) "A quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse blastocyst," *Developmental Biology*, 119(2), pp. 520–531. doi:10.1016/0012-1606(87)90055-8.

Fode, C. *et al.* (1998) "The bHLH Protein NEUROGENIN 2 Is a Determination Factor for Epibranchial Placode–Derived Sensory Neurons," *Neuron*, 20(3), pp. 483–494. doi:10.1016/S0896-6273(00)80989-7.

Fode, C. *et al.* (2000) "A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons," *Genes & Development*, 14(1), pp. 67–80. doi:10.1101/GAD.14.1.67.

Francesconi, M. *et al.* (2019) "Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming," *eLife*, 8. doi:10.7554/ELIFE.41627.

G. Kantzer, C. *et al.* (2017) "Anti-ACSA-2 defines a novel monoclonal antibody for prospective isolation of living neonatal and adult astrocytes," *Glia*, 65(6), pp. 990–1004. doi:10.1002/GLIA.23140.

García-Bellido, A. and de Celis, J.F. (2009) "The complex tale of the achaete-scute complex: A paradigmatic case in the analysis of gene organization and function during development," *Genetics*, 182(3), pp. 631–639. doi:10.1534/GENETICS.109.104083.

Gardner, R.L. (1982) "Investigation of cell lineage and differentiation in the extraembryonic endoderm of the mouse embryo," *Development*, 68(1), pp. 175–198. doi:10.1242/DEV.68.1.175.

Ge, W.P. *et al.* (2012) "Local generation of glia is a major astrocyte source in postnatal cortex," *Nature*, 484(7394), pp. 376–380. doi:10.1038/NATURE10959.

Gilbert, S.F. (2000) *Formation of the Neural Tube*. 6th edn. Sunderland (MA): Sinauer Associates. Available at: https://www.ncbi.nlm.nih.gov/books/NBK10080/ (Accessed: November 25, 2021).

Gomes, F.C.A., Paulin, D. and Neto, V.M. (1999) "Glial fibrillary acidic protein (GFAP): modulation by growth factors and its implication in astrocyte differentiation," *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas*, 32(5), pp. 619–631. doi:10.1590/S0100-879X1999000500016.

Goulding, M., Lumsden, A. and Paquette, A.J. (1994) "Regulation of Pax-3 expression in the dermomyotome and its role in muscle development," *Development*, 120(4), pp. 957–971. doi:10.1242/DEV.120.4.957.

Gouwens, N.W. *et al.* (2020) "Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells," *Cell*, 183(4), pp. 935-953.e19. doi:10.1016/J.CELL.2020.09.057.

Gowan, K. *et al.* (2001) "Crossinhibitory Activities of Ngn1 and Math1 Allow Specification of Distinct Dorsal Interneurons," *Neuron*, 31(2), pp. 219–232. doi:10.1016/S0896-6273(01)00367-1.

Gradwohl, G., Fode, C. and Guillemot, F. (1996) "Restricted Expression of a Novel Murineatonal-Related bHLH Protein in Undifferentiated Neural Precursors," *Developmental Biology*, 180(1), pp. 227–241. doi:10.1006/DBIO.1996.0297.

Gualdi, R. *et al.* (1996) "Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control," *Genes & development*, 10(13), pp. 1670–1682. doi:10.1101/GAD.10.13.1670.

Guillemot, F. *et al.* (1993) "Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons.," *Cell*, 75(3), pp. 463–76. doi:10.1016/0092-8674(93)90381-y.

Guo, L. *et al.* (2019) "Resolving Cell Fate Decisions during Somatic Cell Reprogramming by Single-Cell RNA-Seq," *Molecular Cell*, 73(4), pp. 815-829.e7. doi:10.1016/j.molcel.2019.01.042.

Gurdon, J.B. and Byrne, J.A. (2003) "The first half-century of nuclear transplantation," *Proceedings of the National Academy of Sciences*, 100(14), pp. 8048–8052. doi:10.1073/PNAS.1337135100.

Gurdon, J.B., Elsdale, T.R. and Fischberg, M. (1958) "Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei," *Nature*, 182(4627), pp. 64–65. doi:10.1038/182064a0.

Gurdon, J.B. and Uehlinger, V. (1966) "'Fertile' Intestine Nuclei," *Nature 1966 210:5042*, 210(5042), pp. 1240–1241. doi:10.1038/2101240a0.

Guttenplan, K.A. and Liddelow, S.A. (2019) "Astrocytes and microglia: Models and tools," *The Journal of Experimental Medicine*, 216(1), p. 71. doi:10.1084/JEM.20180200.

Halassa, M.M. *et al.* (2007) "Synaptic Islands Defined by the Territory of a Single Astrocyte," *Journal of Neuroscience*, 27(24), pp. 6473–6477. doi:10.1523/JNEUROSCI.1419-07.2007.

Haque, A. *et al.* (2017) "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.," *Genome medicine*, 9(1), p. 75. doi:10.1186/s13073-017-0467-4.

Haroon, F. *et al.* (2011) "Gp130-dependent astrocytic survival is critical for the control of autoimmune central nervous system inflammation," *Journal of immunology (Baltimore, Md.: 1950)*, 186(11), pp. 6521–6531. doi:10.4049/JIMMUNOL.1001135.

Hasty, P. *et al.* (1993) "Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene," *Nature 1993 364:6437*, 364(6437), pp. 501–506. doi:10.1038/364501a0.

Hatzis, P. and Talianidis, I. (2001) "Regulatory mechanisms controlling human hepatocyte nuclear factor 4alpha gene expression.," *Molecular and cellular biology*, 21(21), pp. 7320–30. doi:10.1128/MCB.21.21.7320-7330.2001.

Heinrich, C. *et al.* (2010) "Directing astroglia from the cerebral cortex into subtype specific functional neurons," *PLoS Biology*, 8(5). doi:10.1371/journal.pbio.1000373.

Heins, N. *et al.* (2002) "Glial cells generate neurons: The role of the transcription factor Pax6," *Nature Neuroscience*, 5(4), pp. 308–315. doi:10.1038/nn828.

Heinz, S. *et al.* (2015) "The selection and function of cell type-specific enhancers," *Nature Reviews Molecular Cell Biology*, 16(3), pp. 144–154. doi:10.1038/NRM3949.

Henke, R.M. *et al.* (2009) "Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube," *Developmental Biology*, 328(2), pp. 529–540. doi:10.1016/J.YDBIO.2009.01.007.

Herrero-Navarro, Á. *et al.* (2021) "Astrocytes and neurons share region-specific transcriptional signatures that confer regional identity to neuronal reprogramming.," *Science advances*, 7(15). doi:10.1126/sciadv.abe8978.

Herrmann, J.E. *et al.* (2008) "STAT3 is a critical regulator of astrogliosis and scar formation after spinal cord injury," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(28), pp. 7231–7243. doi:10.1523/JNEUROSCI.1709-08.2008.

Hirsinger, E. *et al.* (2001) "Notch signalling acts in postmitotic avian myogenic cells to control MyoD activation," *Development*, 128(1), pp. 107–116. doi:10.1242/DEV.128.1.107.

Hobert, O. and Kratsios, P. (2019) "Neuronal identity control by terminal selectors in worms, flies, and chordates," *Current opinion in neurobiology*, 56, pp. 97–105. doi:10.1016/J.CONB.2018.12.006.

Hochedlinger, K. and Jaenisch, R. (2002) "Monoclonal mice generated by nuclear transfer from mature B and T donor cells," *Nature 2002 415:6875*, 415(6875), pp. 1035–1038. doi:10.1038/nature718.

Hochedlinger, K. and Plath, K. (2009) "Epigenetic reprogramming and induced pluripotency," *Development (Cambridge, England)*, 136(4), pp. 509–523. doi:10.1242/DEV.020867.

Hochstim, C. *et al.* (2008) "Identification of Positionally Distinct Astrocyte Subtypes whose Identities Are Specified by a Homeodomain Code," *Cell*, 133(3), pp. 510–522. doi:10.1016/j.cell.2008.02.046.

Holmberg, J. and Perlmann, T. (2012) "Maintaining differentiated cellular identity," *Nature Reviews Genetics 2012 13:6*, 13(6), pp. 429–439. doi:10.1038/nrg3209.

Holtzinger, A. and Evans, T. (2005) "Gata4 regulates the formation of multiple organs," *Development*, 132(17), pp. 4005–4014. doi:10.1242/DEV.01978.

Horton, S. *et al.* (1999) "Correct Coordination of Neuronal Differentiation Events in Ventral Forebrain Requires the bHLH Factor MASH1," *Molecular and Cellular Neuroscience*, 14(4–5), pp. 355–369. doi:10.1006/MCNE.1999.0791.

Horvath, S. (2013) "DNA methylation age of human tissues and cell types.," *Genome biology*, 14(10), p. R115. doi:10.1186/gb-2013-14-10-r115.

Hou, Y. *et al.* (2019) "Ageing as a risk factor for neurodegenerative disease," *Nature Reviews Neurology 2019 15:10*, 15(10), pp. 565–581. doi:10.1038/s41582-019-0244-7.

Hu, X. *et al.* (2019) "Region-Restrict Astrocytes Exhibit Heterogeneous Susceptibility to Neuronal Reprogramming," *Stem Cell Reports*, 12(2), pp. 290–304. doi:10.1016/j.stemcr.2018.12.017.

Huang, H.P. *et al.* (2000) "Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3.," *Molecular and cellular biology*, 20(9), pp. 3292–307. doi:10.1128/MCB.20.9.3292-3307.2000.

Huang, P. *et al.* (2011) "Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors," *Nature*, 475(7356), pp. 386–391. doi:10.1038/nature10116.

Huangfu, D. *et al.* (2008) "Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2," *Nature Biotechnology 2008 26:11*, 26(11), pp. 1269–1275. doi:10.1038/nbt.1502.

Huh, C.J. *et al.* (2016) "Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts," *eLife*, 5(September2016), pp. 1–14. doi:10.7554/eLife.18648.

Huisinga Brent Brower-Toland Sarah C R Elgin, K.L. (2006) "The contradictory definitions of heterochromatin: transcription and silencing," *Chromosoma*, 115, pp. 110–122. doi:10.1007/s00412-006-0052-x.

Hume, D.A. (2000) "Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression," *Blood*, 96(7), pp. 2323–2328. doi:10.1182/BLOOD.V96.7.2323.

Hussein, S.M.I. *et al.* (2014) "Genome-wide characterization of the routes to pluripotency," *Nature*, 516(7530), pp. 198–206. doi:10.1038/NATURE14046.

Imai, H. *et al.* (2020) "Induction of pluripotency in mammalian fibroblasts by cell fusion with mouse embryonic stem cells," *Biochemical and Biophysical Research Communications*, 521(1), pp. 24–30. doi:10.1016/J.BBRC.2019.10.026.

Iwafuchi-Doi, M. and Zaret, K.S. (2014) "Pioneer transcription factors in cell reprogramming," *Genes and Development*. Cold Spring Harbor Laboratory Press, pp. 2679–2692. doi:10.1101/gad.253443.114.

Jan, Y.N. and Jan, L.Y. (1993) "HLH proteins, fly neurogenesis, and vertebrate myogenesis," *Cell*, 75(5), pp. 827–830. doi:10.1016/0092-8674(93)90525-U.

Jessell, T.M. (2000) "Neuronal specification in the spinal cord: inductive signals and transcriptional codes," *Nature Reviews Genetics 2000 1:1*, 1(1), pp. 20–29. doi:10.1038/35049541.

Johansson, C.B. *et al.* (2008) "Extensive fusion of haematopoietic cells with Purkinje neurons in response to chronic inflammation," *Nature Cell Biology 2008 10:5*, 10(5), pp. 575–583. doi:10.1038/ncb1720.

John Lin, C.C. *et al.* (2017) "Identification of diverse astrocyte populations and their malignant analogs," *Nature Neuroscience 2017 20:3*, 20(3), pp. 396–405. doi:10.1038/nn.4493.

Johnson, M.H. and Ziomek, C.A. (1981) "The foundation of two distinct cell lineages within the mouse morula," *Cell*, 24(1), pp. 71–80. doi:10.1016/0092-8674(81)90502-X.

Jostes, B., Walther, C. and Gruss, P. (1990) "The murine paired box gene, Pax7, is expressed specifically during the development of the nervous and muscular system," *Mechanisms of Development*, 33(1), pp. 27–37. doi:10.1016/0925-4773(90)90132-6.

Kablar, B. *et al.* (1999) "Myogenic Determination Occurs Independently in Somites and Limb Buds," *Developmental Biology*, 206(2), pp. 219–231. doi:10.1006/DBIO.1998.9126.

Kamimoto, K., Hoffmann, C.M. and Morris, S.A. (2020) "CellOracle: Dissecting cell identity via network inference and in silico gene perturbation," *bioRxiv*, p. 2020.02.17.947416. doi:10.1101/2020.02.17.947416.

Karamariti, E. *et al.* (2013) "Smooth muscle cells differentiated from reprogrammed embryonic lung fibroblasts through dkk3 signaling are potent for tissue engineering of vascular grafts," *Circulation Research*, 112(11), pp. 1433–1443. doi:10.1161/CIRCRESAHA.111.300415.

Khakh, B.S. (2019) "Astrocyte–Neuron Interactions in the Striatum: Insights on Identity, Form, and Function," *Trends in Neurosciences*, 42(9), pp. 617–630. doi:10.1016/J.TINS.2019.06.003.

Khakh, B.S. and Deneen, B. (2019) "The Emerging Nature of Astrocyte Diversity," *Annual Review of Neuroscience*, 42(1), pp. 187–207. doi:10.1146/annurev-neuro-070918-050443.

Khakh, B.S. and Sofroniew, M. v. (2015) "Diversity of astrocyte functions and phenotypes in neural circuits," *Nature Neuroscience 2015 18:7*, 18(7), pp. 942–952. doi:10.1038/nn.4043.

Kiefer, J.C. and Hauschka, S.D. (2001) "Myf-5 Is Transiently Expressed in Nonmuscle Mesoderm and Exhibits Dynamic Regional Changes within the Presegmented Mesoderm and Somites I–IV," *Developmental Biology*, 232(1), pp. 77–90. doi:10.1006/DBIO.2000.0114.

Kilpinen, H. *et al.* (2017) "Common genetic variation drives molecular heterogeneity in human iPSCs," *Nature 2017 546:7658*, 546(7658), pp. 370–375. doi:10.1038/nature22403.

Kim, E. *et al.* (2020) "Creation of bladder assembloids mimicking tissue regeneration and cancer," *Nature 2020 588:7839*, 588(7839), pp. 664–669. doi:10.1038/s41586-020-3034-x.

Kim, J. *et al.* (2011) "Direct reprogramming of mouse fibroblasts to neural progenitors," *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), pp. 7838–7843. doi:10.1073/pnas.1103113108.

Kim, J., Koo, B.K. and Knoblich, J.A. (2020) "Human organoids: model systems for human biology and medicine," *Nature Reviews Molecular Cell Biology 2020 21:10*, 21(10), pp. 571–584. doi:10.1038/s41580-020-0259-3.

Kim, J.B. *et al.* (2008) "Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors," *Nature 2008 454:7204*, 454(7204), pp. 646–650. doi:10.1038/nature07061.

Kim, Jongpil *et al.* (2017) "Modelling APOE ε3/4 allele-associated sporadic Alzheimer's disease in an induced neuron," *Brain*, 140(8), pp. 2193–2209. doi:10.1093/BRAIN/AWX144.

Kim, Y. *et al.* (2018) "Mitochondrial Aging Defects Emerge in Directly Reprogrammed Human Neurons due to Their Metabolic Profile," *Cell Reports*, 23(9), pp. 2550–2558. doi:10.1016/J.CELREP.2018.04.105.

Kimura, W. *et al.* (2006) "Fate and plasticity of the endoderm in the early chick embryo," *Developmental biology*, 289(2), pp. 283–295. doi:10.1016/J.YDBIO.2005.09.009.

King, T.J. and Briggs, A.R. (1955) "Changes in the nuclei of differentiating gastrula cells, as demonstrated by nuclear transplantation," *Proceedings of the National Academy of Sciences*, 41(5), pp. 321–325. doi:10.1073/PNAS.41.5.321.

Kintner, C. (2002) "Neurogenesis in Embryos and in Adult Neural Stem Cells," *Journal of Neuroscience*, 22(3), pp. 639–643. doi:10.1523/JNEUROSCI.22-03-00639.2002.

Kisiel, M.A. and Klar, A.S. (2019) "Isolation and Culture of Human Dermal Fibroblasts," *Methods in molecular biology (Clifton, N.J.)*, 1993, pp. 71–78. doi:10.1007/978-1-4939-9473-1_6.

Knaupp, A.S. *et al.* (2017) "Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming," *Cell Stem Cell*, 21(6), pp. 834-845.e6. doi:10.1016/j.stem.2017.11.007.

Knust, E. and Campos-Ortega, J.A. (1989) "The molecular genetics of early neurogenesis in Drosophila melanogaster," *BioEssays*, 11(4), pp. 95–100. doi:10.1002/BIES.950110405.

Kornberg, R.D. and Lorch, Y. (1999) "Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome," *Cell*, 98(3), pp. 285–294. doi:10.1016/S0092-8674(00)81958-3.

Kretzschmar, K. and Watt, F.M. (2012) "Lineage tracing," *Cell*, 148(1–2), pp. 33–45. doi:10.1016/J.CELL.2012.01.002.

Kriegstein, A.R. and Götz, M. (2003) "Radial glia diversity: A matter of cell fate," *Glia*, 43(1), pp. 37–43. doi:10.1002/GLIA.10250.

Kuo, C.J. *et al.* (1992) "A transcriptional hierarchy involved in mammalian cell-type specification," *Nature 1992 355:6359*, 355(6359), pp. 457–461. doi:10.1038/355457a0.

Ladewig, J. *et al.* (2012) "Small molecules enable highly efficient neuronal conversion of human fibroblasts," *Nature Methods 2012 9:6*, 9(6), pp. 575–578. doi:10.1038/nmeth.1972.

Laiosa, C. v. *et al.* (2006) "Reprogramming of Committed T Cell Progenitors to Macrophages and Dendritic Cells by C/EBPα and PU.1 Transcription Factors," *Immunity*, 25(5), pp. 731–744. doi:10.1016/J.IMMUNI.2006.09.011.

Lake, B.B. *et al.* (2016) "Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain," *Science (New York, N.Y.)*, 352(6293), p. 1586. doi:10.1126/SCIENCE.AAF1204.

Lancaster, M.A. *et al.* (2013) "Cerebral organoids model human brain development and microcephaly," *Nature 2013 501:7467*, 501(7467), pp. 373–379. doi:10.1038/nature12517.

Lanjakornsiripan, D. *et al.* (2018) "Layer-specific morphological and molecular differences in neocortical astrocytes and their dependence on neuronal layers," *Nature Communications 2018 9:1*, 9(1), pp. 1–15. doi:10.1038/s41467-018-03940-3.

Lapasset, L. *et al.* (2011) "Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state," *Genes & Development*, 25(21), pp. 2248–2253. doi:10.1101/GAD.173922.111.

Lau, H.H. *et al.* (2018) "The molecular functions of hepatocyte nuclear factors – In and beyond the liver," *Journal of Hepatology*, 68(5), pp. 1033–1048. doi:10.1016/j.jhep.2017.11.026.

Lawson, A. and Schoenwolf, G.C. (2003) "Epiblast and primitive-streak origins of the endoderm in the gastrulating chick embryo," *Development (Cambridge, England)*, 130(15), pp. 3491–3501. doi:10.1242/DEV.00579.

Lawson, K.A. (1999) "Fate mapping the mouse embryo.," *International Journal of Developmental Biology*, 43(7), pp. 773–775. doi:10.1387/IJDB.10668985.

Lawson, K.A., Meneses, J.J. and Pedersen, R.A. (1986) "Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer," *Developmental Biology*, 115(2), pp. 325–339. doi:10.1016/0012-1606(86)90253-8.

Lawson, K.A. and Pedersen, R.A. (1992) "Clonal analysis of cell fate during gastrulation and early neurulation in the mouse," *Ciba Foundation symposium*, 165. doi:10.1002/9780470514221.CH2.

Lazzaro, D. *et al.* (1992) "LFB1 and LFB3 homeoproteins are sequentially expressed during kidney development.," *Development (Cambridge, England)*, 114(2), pp. 469–79. doi:10.1242/dev.114.2.469.

Lee, C.S. *et al.* (2005) "The initiation of liver development is dependent on Foxa transcription factors," *Nature 2005 435:7044*, 435(7044), pp. 944–947. doi:10.1038/nature03649.

Lee, J.E. *et al.* (1995) "Conversion of Xenopus Ectoderm into Neurons by NeuroD, a Basic Helix-Loop-Helix Protein," *Science*, 268(5212), pp. 836–844. doi:10.1126/SCIENCE.7754368.

Lee, J.E. (1997) "Basic helix-loop-helix genes in neural development," *Current Opinion in Neurobiology*, 7(1), pp. 13–20. doi:10.1016/S0959-4388(97)80115-8.

Lee, J.E. *et al.* (2020) "Neuroprotective Effects of Cryptotanshinone in a Direct Reprogramming Model of Parkinson's Disease," *Molecules 2020, Vol. 25, Page 3602*, 25(16), p. 3602. doi:10.3390/MOLECULES25163602.

Lee, Q.Y. *et al.* (2020) "Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes," *Nature Cell Biology*, 22(4), pp. 401–411. doi:10.1038/s41556-020-0490-3.

Li, D. *et al.* (2017) "Chromatin Accessibility Dynamics during iPSC Reprogramming," *Cell Stem Cell*, 21(6), pp. 819-833.e6. doi:10.1016/j.stem.2017.10.012.

Li, L. *et al.* (2008) "Protective role of reactive astrocytes in brain ischemia," *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 28(3), pp. 468–481. doi:10.1038/SJ.JCBFM.9600546.

Li, R. *et al.* (2010) "A mesenchymal-to-Epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts," *Cell Stem Cell*, 7(1), pp. 51–63. doi:10.1016/j.stem.2010.04.014.

Li, Z. *et al.* (2009) "Foxa1 and Foxa2 regulate bile duct development in mice," *The Journal of Clinical Investigation*, 119(6), pp. 1537–1545. doi:10.1172/JCI38201.

Lim, K.T. *et al.* (2016) "Small Molecules Facilitate Single Factor-Mediated Hepatic Reprogramming," *Cell Reports*, 15(4), pp. 814–829. doi:10.1016/j.celrep.2016.03.071.

Liu, M. *et al.* (2000) "Loss of BETA2/NeuroD leads to malformation of the dentate gyrus and epilepsy," *Proceedings of the National Academy of Sciences*, 97(2), pp. 865–870. doi:10.1073/PNAS.97.2.865.

Liu, M.L., Zang, T. and Zhang, C.L. (2016) "Direct Lineage Reprogramming Reveals Disease-Specific Phenotypes of Motor Neurons from Human ALS Patients," *Cell Reports*, 14(1), pp. 115–128. doi:10.1016/J.CELREP.2015.12.018.

Liu, X. *et al.* (2020) "Reprogramming roadmap reveals route to human induced trophoblast stem cells," *Nature 2020 586:7827*, 586(7827), pp. 101–107. doi:10.1038/s41586-020-2734-6.

Liu, Y. *et al.* (2014) "Direct Reprogramming of Huntington's Disease Patient Fibroblasts into Neuron-Like Cells Leads to Abnormal Neurite Outgrowth, Increased Cell Death, and Aggregate Formation," *PLOS ONE*, 9(10), p. e109621. doi:10.1371/JOURNAL.PONE.0109621.

Longo, S.K. *et al.* (2021) "Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics," *Nature Reviews Genetics 2021 22:10*, 22(10), pp. 627–644. doi:10.1038/s41576-021-00370-8.

Luger, K. *et al.* (1997) "Crystal structure of the nucleosome core particle at 2.8 Å resolution," *Nature 1997 389:6648*, 389(6648), pp. 251–260. doi:10.1038/38444.

Ma, Q. *et al.* (1998) "Neurogenin1 is essential for the determination of neuronal precursors for proximal cranial sensory ganglia.," *Neuron*, 20(3), pp. 469–82. doi:10.1016/s0896-6273(00)80988-5.

Ma, Q. *et al.* (1999) "Neurogenin1 and neurogenin2 control two distinct waves of neurogenesis in developing dorsal root ganglia.," *Genes & development*, 13(13), pp. 1717–28. doi:10.1101/gad.13.13.1717.

Ma, Q., Kintner, C. and Anderson, D.J. (1996) "Identification of neurogenin, a vertebrate neuronal determination gene.," *Cell*, 87(1), pp. 43–52. doi:10.1016/s0092-8674(00)81321-5.

Macauley, S.L., Pekny, M. and Sands, M.S. (2011) "The role of attenuated astrocyte activation in infantile neuronal ceroid lipofuscinosis," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(43), pp. 15575–15585. doi:10.1523/JNEUROSCI.3579-11.2011.

Maherali, N. *et al.* (2007) "Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution," *Cell Stem Cell*, 1(1), pp. 55–70. doi:10.1016/j.stem.2007.05.014.

Malinowski, A.R. and Fisher, A.G. (2016) "Reprogramming of Somatic Cells Towards Pluripotency by Cell Fusion," *Methods in Molecular Biology*, 1480, pp. 289–299. doi:10.1007/978-1-4939-6380-5_25.

Mancinelli, S. and Lodato, S. (2018) "Decoding neuronal diversity in the developing cerebral cortex: from single cells to functional networks," *Current Opinion in Neurobiology*, 53, pp. 146–155. doi:10.1016/J.CONB.2018.08.001.

Margariti, A. *et al.* (2012) "Direct reprogramming of fibroblasts into endothelial cells capable of angiogenesis and reendothelialization in tissue-engineered vessels.," *Proceedings of the National Academy of Sciences of the United States of America*, 109(34), pp. 13793–8. doi:10.1073/pnas.1205526109.

Marro, S. *et al.* (2011) "Direct lineage conversion of terminally differentiated hepatocytes to functional neurons," *Cell stem cell*, 9(4), p. 374. doi:10.1016/J.STEM.2011.09.002.

Marx, V. (2021) "Method of the Year: spatially resolved transcriptomics," *Nature Methods 2021 18:1*, 18(1), pp. 9–14. doi:10.1038/s41592-020-01033-y.

Massari, M.E. and Murre, C. (2000) "Helix-Loop-Helix Proteins: Regulators of Transcription in Eucaryotic Organisms," *Molecular and Cellular Biology*, 20(2), pp. 429–440. doi:10.1128/MCB.20.2.429-440.2000.

Masserdotti, G. *et al.* (2015) "Transcriptional Mechanisms of Proneural Factors and REST in Regulating Neuronal Reprogramming of Astrocytes," *Cell Stem Cell*, 17(1), pp. 74–88. doi:10.1016/j.stem.2015.05.014.

Mattugini, N. *et al.* (2019) "Inducing Different Neuronal Subtypes from Astrocytes in the Injured Mouse Cerebral Cortex," *Neuron*, 103(6), pp. 1086-1095.e5. doi:10.1016/J.NEURON.2019.08.009.

Mayran, A. and Drouin, J. (2018) "Pioneer transcription factors shape the epigenetic landscape," *Journal of Biological Chemistry*, 293(36), pp. 13795–13804. doi:10.1074/JBC.R117.001232.

Maza, I. *et al.* (2015) "Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors," *Nature Biotechnology 2015 33:7*, 33(7), pp. 769–774. doi:10.1038/nbt.3270.

McKenna, A. and Gagnon, J.A. (2019) "Recording development with single cell dynamic lineage tracing," *Development (Cambridge)*, 146(12). doi:10.1242/DEV.169730.

Mertens, J. *et al.* (2015) "Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects," *Cell Stem Cell*, 17(6), pp. 705–718. doi:10.1016/j.stem.2015.09.001.

Mikkelsen, T.S. *et al.* (2008) "Dissecting direct reprogramming through integrative genomic analysis," *Nature*, 454(7200), pp. 49–55. doi:10.1038/NATURE07056.

Miller, F.D. and Gauthier, A.S. (2007) "Timing Is Everything: Making Neurons versus Glia in the Developing Cortex," *Neuron*, 54(3), pp. 357–369. doi:10.1016/J.NEURON.2007.04.019.

Miller, J.D. *et al.* (2013) "Human iPSC-Based Modeling of Late-Onset Disease via Progerin-Induced Aging," *Cell Stem Cell*, 13(6), pp. 691–705. doi:10.1016/J.STEM.2013.11.006.

Misra, K. *et al.* (2014) "Asymmetric activation of Dll4-Notch signaling by Foxn4 and proneural factors activates BMP/TGFβ signaling to specify V2b interneurons in the spinal cord," *Development*, 141(1), pp. 187–198. doi:10.1242/DEV.092536.

Mitchell, R. *et al.* (2014) "Molecular Evidence for OCT4-Induced Plasticity in Adult Human Fibroblasts Required for Direct Cell Fate Conversion to Lineage Specific Progenitors," *Stem Cells*, 32(8), pp. 2178–2187. doi:10.1002/STEM.1721.

Miyata, T., Maeda, T. and Lee, J.E. (1999) "NeuroD is required for differentiation of the granule cells in the cerebellum and hippocampus," *Genes & Development*, 13(13), p. 1647. doi:10.1101/GAD.13.13.1647.

Molofsky, A.V. and Deneen, B. (2015) "Astrocyte development: A Guide for the Perplexed," *Glia*, 63(8), pp. 1320–1329. doi:10.1002/glia.22836.

Morel, L., Chiang, Ming Sum R, *et al.* (2017) "Molecular and Functional Properties of Regional Astrocytes in the Adult Brain.," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 37(36), pp. 8706–8717. doi:10.1523/JNEUROSCI.3956-16.2017.

Morel, L., Chiang, Ming Sum R., *et al.* (2017) "Molecular and Functional Properties of Regional Astrocytes in the Adult Brain," *The Journal of Neuroscience*, pp. 3956–16. doi:10.1523/JNEUROSCI.3956-16.2017.

Moris, N., Pina, C. and Arias, A.M. (2016) "Transition states and cell fate decisions in epigenetic landscapes," *Nature Reviews Genetics 2016 17:11*, 17(11), pp. 693–703. doi:10.1038/nrg.2016.98.

Morris, S.A. *et al.* (2014) "Dissecting Engineered Cell Types and Enhancing Cell Fate Conversion via CellNet," *Cell*, 158, pp. 889–902. doi:10.1016/j.cell.2014.07.021.

Morris, S.A. (2016) "Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks," *Development*, 143(15), pp. 2696–2705. doi:10.1242/dev.138263.

Morris, Samantha A (2019) "The evolving concept of cell identity in the single cell era." doi:10.1242/dev.169748.

Morris, Samantha A. (2019) "The evolving concept of cell identity in the single cell era," *Development (Cambridge)*, 146(12). doi:10.1242/dev.169748.

Morris, S.A. and Daley, G.Q. (2013) "A blueprint for engineering cell fate: current technologies to reprogram cell identity," *Cell Research 2013 23:1*, 23(1), pp. 33–48. doi:10.1038/cr.2013.1.

Mortazavi, A. *et al.* (2008) "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods*, 5(7), pp. 621–628. doi:10.1038/nmeth.1226.

Murre, C. (2019) "Helix–loop–helix proteins and the advent of cellular diversity: 30 years of discovery," *Genes & Development*, 33(1–2), pp. 6–25. doi:10.1101/GAD.320663.118.

Murre, C., McCaw, P.S. and Baltimore, D. (1989) "A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins," *Cell*, 56(5), pp. 777–783. doi:10.1016/0092-8674(89)90682-X.

Myer, D.J. *et al.* (2006) "Essential protective roles of reactive astrocytes in traumatic brain injury.," *Brain : a journal of neurology*, 129(Pt 10), pp. 2761–72. doi:10.1093/brain/awl165.

Nabeshima, Y. *et al.* (1993) "Myogenin gene disruption results in perinatal lethality because of severe muscle defect," *Nature 1993 364:6437*, 364(6437), pp. 532–535. doi:10.1038/364532a0.

Nawashiro, H. *et al.* (1998) "Mice lacking GFAP are hypersensitive to traumatic cerebrospinal injury," *Neuroreport*, 9(8), pp. 1691–1696. doi:10.1097/00001756-199806010-00004.

Nefzger, C.M. *et al.* (2017) "Cell Type of Origin Dictates the Route to Pluripotency," *Cell Reports*, 21(10), pp. 2649–2660. doi:10.1016/j.celrep.2017.11.029.

Nemajerova, A. *et al.* (2012) "Two-factor reprogramming of somatic cells to pluripotent stem cells reveals partial functional redundancy of Sox2 and Klf4," *Cell Death and Differentiation*, 19(8), pp. 1268–1276. doi:10.1038/cdd.2012.45.

Nieto, M. *et al.* (2001) "Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors.," *Neuron*, 29(2), pp. 401–13. doi:10.1016/s0896-6273(01)00214-8.

Odom, D.T. *et al.* (2004) "Control of Pancreas and Liver Gene Expression by HNF Transcription Factors," *Science (New York, N.Y.)*, 303(5662), p. 1378. doi:10.1126/SCIENCE.1089769.

Ohlig, S. *et al.* (2021) "Molecular diversity of diencephalic astrocytes reveals adult astrogenesis regulated by Smad4," *The EMBO Journal*, 40(21), p. e107532. doi:10.15252/EMBJ.2020107532.

Olson, J.M. *et al.* (2001) "NeuroD2 Is Necessary for Development and Survival of Central Nervous System Neurons," *Developmental Biology*, 234(1), pp. 174–187. doi:10.1006/DBIO.2001.0245.

Orge, I.D. *et al.* (2020) "Phenotype instability of hepatocyte-like cells produced by direct reprogramming of mesenchymal stromal cells," *Stem cell research & therapy*, 11(1). doi:10.1186/S13287-020-01665-Z.

Özel, M.N. *et al.* (2020) "Neuronal diversity and convergence in a visual system developmental atlas," *Nature 2020 589:7840*, 589(7840), pp. 88–95. doi:10.1038/s41586-020-2879-3.

Pang, Z.P. *et al.* (2011) "Induction of human neuronal cells by defined transcription factors," *Nature*. Nature Publishing Group, pp. 220–223. doi:10.1038/nature10202.

Parenti, A. *et al.* (2016) "OSKM Induce Extraembryonic Endoderm Stem Cells in Parallel to Induced Pluripotent Stem Cells," *Stem Cell Reports*, 6(4), pp. 447–455. doi:10.1016/J.STEMCR.2016.02.003.

Park, I.H. *et al.* (2007) "Reprogramming of human somatic cells to pluripotency with defined factors," *Nature 2007 451:7175*, 451(7175), pp. 141–146. doi:10.1038/nature06534.

Parker, M.H., Seale, P. and Rudnicki, M.A. (2003) "Looking back to the embryo: defining transcriptional networks in adult myogenesis," *Nature Reviews Genetics 2003 4:7*, 4(7), pp. 497–507. doi:10.1038/nrg1109.

Pashos, E.E. *et al.* (2017) "Large, Diverse Population Cohorts of hiPSCs and Derived Hepatocyte-like Cells Reveal Functional Genetic Variation at Blood Lipid-Associated Loci," *Cell Stem Cell*, 20(4), pp. 558-570.e10. doi:10.1016/J.STEM.2017.03.017.

Patapoutian, A. *et al.* (1995) "Disruption of the mouse MRF4 gene identifies multiple waves of myogenesis in the myotome," *Development*, 121(10), pp. 3347–3358. doi:10.1242/DEV.121.10.3347.

Patthey, C. and Gunhaga, L. (2014) "Signaling pathways regulating ectodermal cell fate choices," *Experimental Cell Research*, 321(1), pp. 11–16. doi:10.1016/J.YEXCR.2013.08.002.

Pedersen, R.A., Wu, K. and Bałakier, H. (1986) "Origin of the inner cell mass in mouse embryos: Cell lineage analysis by microinjection," *Developmental Biology*, 117(2), pp. 581–595. doi:10.1016/0012-1606(86)90327-1.

Perron, M. *et al.* (1999) "X-ngnr-1 and Xath3 promote ectopic expression of sensory neuron markers in the neurula ectoderm and have distinct inducing properties in the retina," *Proceedings of the National Academy of Sciences*, 96(26), pp. 14996–15001. doi:10.1073/PNAS.96.26.14996.

Peskova, L. *et al.* (2019) "Oct4-mediated reprogramming induces embryonic-like microRNA expression signatures in human fibroblasts," *Scientific Reports*, 9(1), pp. 1–13. doi:10.1038/s41598-019-52294-3.

Pfisterer, U. *et al.* (2011) "Direct conversion of human fibroblasts to dopaminergic neurons," *Proceedings of the National Academy of Sciences*, 108(25), pp. 10343–10348. doi:10.1073/pnas.1105135108.

Pfisterer, Ulrich *et al.* (2011) "Efficient induction of functional neurons from adult human fibroblasts.," *Cell cycle (Georgetown, Tex.)*, 10(19), pp. 3311–6. doi:10.4161/cc.10.19.17584.

Philippidou, P. and Dasen, J.S. (2013) "Hox Genes: Choreographers in Neural Development, Architects of Circuit Organization," *Neuron*, 80(1), pp. 12–34. doi:10.1016/J.NEURON.2013.09.020.

Pircs, K. *et al.* (2018) "Huntingtin Aggregation Impairs Autophagy, Leading to Argonaute-2 Accumulation and Global MicroRNA Dysregulation.," *Cell reports*, 24(6), pp. 1397–1406. doi:10.1016/j.celrep.2018.07.017.

Pircs, K. *et al.* (2021) "Distinct subcellular autophagy impairments in induced neurons from Huntington's disease patients," *Brain*, 139(4), pp. 16–17. doi:10.1093/brain/awab473.

Plikus, M. v. *et al.* (2021) "Fibroblasts: Origins, definitions, and functions in health and disease," *Cell*, 184(15), pp. 3852–3872. doi:10.1016/J.CELL.2021.06.024.

Plusa, B. *et al.* (2008) "Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst," *Development*, 135(18), pp. 3081–3091. doi:10.1242/DEV.021519.

Pournasr, B., Asghari-Vostikolaee, M.H. and Baharvand, H. (2015) "Transcription factor-mediated reprograming of fibroblasts to hepatocyte-like cells," *European Journal of Cell Biology*, 94(12), pp. 603–610. doi:10.1016/J.EJCB.2015.10.003.

Pourquié, O. (2001) "Vertebrate somitogenesis.," *Annual review of cell and developmental biology*, 17, pp. 311–50. doi:10.1146/annurev.cellbio.17.1.311.

Powell, A.E. *et al.* (2011) "Fusion between Intestinal Epithelial Cells and Macrophages in a Cancer Context Results in Nuclear Reprogramming," *Cancer Research*, 71(4), pp. 1497–1505. doi:10.1158/0008-5472.CAN-10-3223.

Pownall, M.E., Gustafsson, M.K. and Emerson, C.P. (2002) "Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos.," *Annual review of cell and developmental biology*, 18, pp. 747–83. doi:10.1146/annurev.cellbio.18.012502.105758.

Pralong, D., Trounson, A.O. and Verma, P.J. (2006) "Cell fusion for reprogramming pluripotency: toward elimination of the pluripotent genome.," *Stem cell reviews*, 2(4), pp. 331–40. doi:10.1007/BF02698060.

Protze, S. *et al.* (2012) "A new approach to transcription factor screening for reprogramming of fibroblasts to cardiomyocyte-like cells," *Journal of Molecular and Cellular Cardiology*, 53(3), pp. 323–332. doi:10.1016/J.YJMCC.2012.04.010.

Qu, H. *et al.* (2000) "13C MR Spectroscopy Study of Lactate as Substrate for Rat Brain," *Developmental Neuroscience*, 22(5–6), pp. 429–436. doi:10.1159/000017472.

Quadrato, G. *et al.* (2017) "Cell diversity and network dynamics in photosensitive human brain organoids," *Nature*, 545(7652), pp. 48–53. doi:10.1038/NATURE22047.

Raj, B. *et al.* (2018) "Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain," *Nature Biotechnology*, 36(5), pp. 442–450. doi:10.1038/NBT.4103.

Ransohoff, R.M. (2018) "All (animal) models (of neurodegeneration) are wrong. Are they also useful?," *The Journal of Experimental Medicine*, 215(12), p. 2955. doi:10.1084/JEM.20182042.

Rao, Z. *et al.* (2021) "Molecular Mechanisms Underlying Ascl1-Mediated Astrocyte-to-Neuron Conversion," *Stem Cell Reports*, 16(3), pp. 534–547. doi:10.1016/J.STEMCR.2021.01.006.

Rawlings, T.M. *et al.* (2021) "Modelling the impact of decidual senescence on embryo implantation in human endometrial assembloids," *eLife*, 10. doi:10.7554/ELIFE.69603.

Rizvi, A.Z. *et al.* (2006) "Bone marrow-derived cells fuse with normal and transformed intestinal stem cells," *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), pp. 6321–6325. doi:10.1073/PNAS.0508593103.

Rodriques, S.G. *et al.* (2019) "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.," *Science (New York, N.Y.)*, 363(6434), pp. 1463–1467. doi:10.1126/science.aaw1219.

Rombaut, M. *et al.* (2021) "Direct reprogramming of somatic cells into induced hepatocytes: Cracking the Enigma code.," *Journal of hepatology*, 75(3), pp. 690–705. doi:10.1016/j.jhep.2021.04.048.

Rothstein, J.D. *et al.* (1994) "Localization of neuronal and glial glutamate transporters," *Neuron*, 13(3), pp. 713–725. doi:10.1016/0896-6273(94)90038-8.

Rothstein, J.D. *et al.* (1996) "Knockout of glutamate transporters reveals a major role for astroglial transport in excitotoxicity and clearance of glutamate," *Neuron*, 16(3), pp. 675–686. doi:10.1016/S0896-6273(00)80086-0.

Rouach, N. *et al.* (2008) "Astroglial metabolic networks sustain hippocampal synaptic transmission," *Science (New York, N.Y.)*, 322(5907), pp. 1551–1555. doi:10.1126/SCIENCE.1164022.

Rowe, R.G. and Daley, G.Q. (2019) "Induced pluripotent stem cells in disease modelling and drug discovery," *Nature Reviews Genetics 2019 20:7*, 20(7), pp. 377–388. doi:10.1038/s41576-019-0100-z.

Rowitch, D.H. and Kriegstein, A.R. (2010) "Developmental genetics of vertebrate glial–cell specification," *Nature 2010 468:7321*, 468(7321), pp. 214–222. doi:10.1038/nature09611.

Rudnicki, M.A. *et al.* (1992) "Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development," *Cell*, 71(3), pp. 383–390. doi:10.1016/0092-8674(92)90508-A.

Rudnicki, M.A. *et al.* (1993) "MyoD or Myf-5 is required for the formation of skeletal muscle," *Cell*, 75(7), pp. 1351–1359. doi:10.1016/0092-8674(93)90621-V.

Russo, G.L. *et al.* (2021) "CRISPR-Mediated Induction of Neuron-Enriched Mitochondrial Proteins Boosts Direct Glia-to-Neuron Conversion," *Cell stem cell*, 28(3), pp. 524-534.e7. doi:10.1016/J.STEM.2020.10.015.

Samavarchi-Tehrani, P. *et al.* (2010) "Functional genomics reveals a BMP-Driven mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming," *Cell Stem Cell*, 7(1), pp. 64–77. doi:10.1016/j.stem.2010.04.015.

Sansom, S.N. *et al.* (2009) "The Level of the Transcription Factor Pax6 Is Essential for Controlling the Balance between Neural Stem Cell Self-Renewal and Neurogenesis," *PLOS Genetics*, 5(6), p. e1000511. doi:10.1371/JOURNAL.PGEN.1000511.

Santello, M., Toni, N. and Volterra, A. (2019) "Astrocyte function from information processing to cognition and cognitive impairment," *Nature Neuroscience 2019 22:2*, 22(2), pp. 154–166. doi:10.1038/s41593-018-0325-8.

Saunders, A. *et al.* (2018) "Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain," *Cell*, 174(4), pp. 1015-1030.e16. doi:10.1016/j.cell.2018.07.028.

Scardigli, R. *et al.* (2001) "Crossregulation between Neurogenin2 and pathways specifying neuronal identity in the spinal cord.," *Neuron*, 31(2), pp. 203–17. doi:10.1016/s0896-6273(01)00358-0.

Schalch, T. *et al.* (2005) "X-ray structure of a tetranucleosome and its implications for the chromatin fibre," *Nature 2005 436:7047*, 436(7047), pp. 138–141. doi:10.1038/nature03686.

Schiebinger, G. *et al.* (2019) "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming," *Cell*, 176(4), pp. 928-943.e22. doi:10.1016/J.CELL.2019.01.006.

Schwab, M.H. *et al.* (2000) "Neuronal Basic Helix-Loop-Helix Proteins (NEX and BETA2/Neuro D) Regulate Terminal Granule Cell Differentiation in the Hippocampus," *Journal of Neuroscience*, 20(10), pp. 3714–3724. doi:10.1523/JNEUROSCI.20-10-03714.2000.

Schwarz, B.A. *et al.* (2018) "Prospective Isolation of Poised iPSC Intermediates Reveals Principles of Cellular Reprogramming," *Cell Stem Cell*, 23(2), pp. 289-305.e5. doi:10.1016/j.stem.2018.06.013.

Schweitzer, J.S. *et al.* (2020) "Personalized iPSC-Derived Dopamine Progenitor Cells for Parkinson's Disease.," *The New England journal of medicine*, 382(20), pp. 1926–1932. doi:10.1056/NEJMoa1915872.

Sekiya, S. and Suzuki, A. (2011) "Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors," *Nature*, 475. doi:10.1038/nature10263.

Sharma, P. *et al.* (2019) "Oct4 mediates Müller glia reprogramming and cell cycle exit during retina regeneration in zebrafish," *Life Science Alliance*, 2(5). doi:10.26508/lsa.201900548.

Silver, J. and Miller, J.H. (2004) "Regeneration beyond the glial scar," *Nature reviews. Neuroscience*, 5(2), pp. 146–156. doi:10.1038/NRN1326.

Simpson, P. (1990) "Notch and the choice of cell fate in Drosophila neuroepithelium," *Trends in Genetics*, 6(C), pp. 343–345. doi:10.1016/0168-9525(90)90260-D.

Siqueira, M. *et al.* (2018) "Radial Glia Cells Control Angiogenesis in the Developing Cerebral Cortex Through TGF-β1 Signaling.," *Molecular neurobiology*, 55(5), pp. 3660–3675. doi:10.1007/s12035-017-0557-8.

Si-Tayeb, K., Lemaigre, F.P. and Duncan, S.A. (2010) "Organogenesis and Development of the Liver," *Developmental Cell*, 18(2), pp. 175–189. doi:10.1016/J.DEVCEL.2010.01.011.

Smith, D.K. *et al.* (2016) "Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming," *Stem Cell Reports*, 7(5), pp. 955–969. doi:10.1016/j.stemcr.2016.09.013.

Sofroniew, M. v. (2009) "Molecular dissection of reactive astrogliosis and glial scar formation," *Trends in neurosciences*, 32(12), p. 638. doi:10.1016/J.TINS.2009.08.002.

Sofroniew, M. v. (2015) "Astrogliosis," *Cold Spring Harbor Perspectives in Biology*, 7(2), p. a020420. doi:10.1101/CSHPERSPECT.A020420.

Sofroniew, M. v. and Vinters, H. v. (2010) "Astrocytes: biology and pathology," *Acta neuropathologica*, 119(1), pp. 7–35. doi:10.1007/S00401-009-0619-8.

Sotelo-Hitschfeld, T. *et al.* (2015) "Channel-Mediated Lactate Release by K+-Stimulated Astrocytes," *Journal of Neuroscience*, 35(10), pp. 4168–4178. doi:10.1523/JNEUROSCI.5036-14.2015.

Soufi, A., Donahue, G. and Zaret, K.S. (2012) "Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome," *Cell*, 151(5), pp. 994–1004. doi:10.1016/J.CELL.2012.09.045.

de Souza, N. (2018) "Organoids," *Nature Methods*, 15(1), pp. 23–23. doi:10.1038/nmeth.4576.

Spees, J.L. *et al.* (2003) "Differentiation, cell fusion, and nuclear fusion during ex vivo repair of epithelium by human adult stem cells from bone marrow stroma," *Proceedings of the National Academy of Sciences*, 100(5), pp. 2397–2402. doi:10.1073/PNAS.0437997100.

Spicer, D.B. *et al.* (1996) "Inhibition of Myogenic bHLH and MEF2 Transcription Factors by the bHLH Protein Twist," *Science*, 272(5267), pp. 1476–1480. doi:10.1126/SCIENCE.272.5267.1476.

Spitz, F. and Furlong, E.E.M. (2012) "Transcription factors: From enhancer binding to developmental control," *Nature Reviews Genetics*, 13(9), pp. 613–626. doi:10.1038/NRG3207.

Sridharan, R. *et al.* (2009) "Role of the Murine Reprogramming Factors in the Induction of Pluripotency," *Cell*, 136(2), pp. 364–377. doi:10.1016/j.cell.2009.01.001.

Stadtfeld, M. *et al.* (2008) "Defining Molecular Cornerstones during Fibroblast to iPS Cell Reprogramming in Mouse," *Cell Stem Cell*, 2(3), pp. 230–240. doi:10.1016/j.stem.2008.02.001.

Ståhl, P.L. *et al.* (2016) "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.," *Science (New York, N.Y.)*, 353(6294), pp. 78–82. doi:10.1126/science.aaf2403.

Stickels, R.R. *et al.* (2020) "Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2," *Nature Biotechnology 2020 39:3*, 39(3), pp. 313–319. doi:10.1038/s41587-020-0739-1.

Sulston, J.E. *et al.* (1983) "The embryonic cell lineage of the nematode Caenorhabditis elegans," *Developmental Biology*, 100(1), pp. 64–119. doi:10.1016/0012-1606(83)90201-4.

Summerbell, D. and Rigby, P.W.J. (1999) "Transcriptional Regulation during Somitogenesis," *Current Topics in Developmental Biology*, 48(C), pp. 301–318. doi:10.1016/S0070-2153(08)60760-7.

Svensson, V., Vento-Tormo, R. and Teichmann, S.A. (2018) "Exponential scaling of single-cell RNA-seq in the past decade," *Nature Protocols 2018 13:4*, 13(4), pp. 599–604. doi:10.1038/nprot.2017.149.

Takahashi, K. *et al.* (2007) "Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors," *Cell*, 131(5), pp. 861–872. doi:10.1016/J.CELL.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," *Cell*, 126(4), pp. 663–676. doi:10.1016/j.cell.2006.07.024.

Tang, Y. *et al.* (2017) "Direct Reprogramming Rather than iPSC-Based Reprogramming Maintains Aging Hallmarks in Human Motor Neurons.," *Frontiers in molecular neuroscience*, 10, p. 359. doi:10.3389/fnmol.2017.00359.

Taylor, S.M. and Jones, P.A. (1979) "Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine.," *Cell*, 17(4), pp. 771–9. doi:10.1016/0092-8674(79)90317-9.

Tran, K.A. *et al.* (2019) "Defining Reprogramming Checkpoints from Single-Cell Analyses of Induced Pluripotency," *Cell Reports*, 27(6), pp. 1726-1741.e5. doi:10.1016/j.celrep.2019.04.056.

Treutlein, B. *et al.* (2016) "Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq.," *Nature*, 534(7607), pp. 391–5. doi:10.1038/nature18323.

Tsai, H.-H. *et al.* (2012) "Regional Astrocyte Allocation Regulates CNS Synaptogenesis and Repair," *Science*, 337(6092), pp. 358–362. doi:10.1126/science.1222381.

di Tullio, A. and Graf, T. (2012) "C/EBPα bypasses cell cycle-dependency during immune cell transdifferentiation.," *Cell cycle (Georgetown, Tex.)*, 11(14), pp. 2739–46. doi:10.4161/cc.21119.

Various Authors (2017) "What Is Your Conceptual Definition of 'Cell Type' in the Context of a Mature Organism?," *Cell systems*, 4(3), pp. 255–259. doi:10.1016/J.CELS.2017.03.006.

Vassilopoulos, G., Wang, P.R. and Russell, D.W. (2003) "Transplanted bone marrow regenerates liver by cell fusion," *Nature 2003 422:6934*, 422(6934), pp. 901–904. doi:10.1038/nature01539.

Velasco, S. *et al.* (2019) "Individual brain organoids reproducibly form cell diversity of the human cerebral cortex," *Nature 2019 570:7762*, 570(7762), pp. 523–527. doi:10.1038/s41586-019-1289-x.

Verkhratsky, A. and Nedergaard, M. (2014) "Astroglial cradle in the life of the synapse," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1654). doi:10.1098/RSTB.2013.0595.

Verkhratsky, A. and Nedergaard, M. (2018) "Physiology of Astroglia," *Physiological Reviews*, 98(1), p. 239. doi:10.1152/PHYSREV.00042.2016.

Vickovic, S. *et al.* (2019) "High-definition spatial transcriptomics for in situ tissue profiling," *Nature Methods 2019 16:10*, 16(10), pp. 987–990. doi:10.1038/s41592-019-0548-y.

Victor, M.B. *et al.* (2018) "Striatal neurons directly converted from Huntington's disease patient fibroblasts recapitulate age-associated disease phenotypes," *Nature neuroscience*, 21(3), pp. 341–352. doi:10.1038/S41593-018-0075-7.

Vierbuchen, T. *et al.* (2010) "Direct conversion of fibroblasts to functional neurons by defined factors," *Nature*, 463(7284), pp. 1035–1041. doi:10.1038/nature08797.

Vogt, N. (2021) "Assembloids," *Nature Methods 2021 18:1*, 18(1), pp. 27–27. doi:10.1038/s41592-020-01026-x.

Volterra, A. and Meldolesi, J. (2005) "Astrocytes, from brain glue to communication elements: the revolution continues," *Nature reviews. Neuroscience*, 6(8), pp. 626–640. doi:10.1038/NRN1722.

Voskuhl, R.R. *et al.* (2009) "Reactive astrocytes form scar-like perivascular barriers to leukocytes during adaptive immune inflammation of the CNS," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(37), pp. 11511–11522. doi:10.1523/JNEUROSCI.1514-09.2009.

Waddington, C.H. (1957) *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*. London: Allen & Unwin.

Wagner, D.E. *et al.* (2018) "Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo," *Science*, 360(6392), pp. 981–987. doi:10.1126/SCIENCE.AAR4362.

Wakayama, T. *et al.* (1998) "Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei," *Nature 1998 394:6691*, 394(6691), pp. 369–374. doi:10.1038/28615.

Wakayama, T. *et al.* (1999) "Mice cloned from embryonic stem cells," *Proceedings of the National Academy of Sciences*, 96(26), pp. 14984–14989. doi:10.1073/PNAS.96.26.14984.

Wang, H. *et al.* (2021) "Direct cell reprogramming: approaches, mechanisms and progress," *Nature Reviews Molecular Cell Biology 2021 22:6*, 22(6), pp. 410–424. doi:10.1038/s41580-021-00335-z.

Wang, X. *et al.* (2003) "Cell fusion is the principal source of bone-marrow-derived hepatocytes," *Nature 2003 422:6934*, 422(6934), pp. 897–901. doi:10.1038/nature01531.

Wanner, I.B. *et al.* (2013) "Glial scar borders are formed by newly proliferated, elongated astrocytes that interact to corral inflammatory and fibrotic cells via STAT3-dependent mechanisms after spinal cord injury," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(31), pp. 12870–12886. doi:10.1523/JNEUROSCI.2121-13.2013.

Wapinski, O.L. *et al.* (2017) "Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons.," *Cell reports*, 20(13), pp. 3236–3247. doi:10.1016/j.celrep.2017.09.011.

Warren, C. (2017) "Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease," *Cell Stem Cell*, 20, pp. 547–557.

Watanabe, S. *et al.* (2011) "MyoD gene suppression by Oct4 is required for reprogramming in myoblasts to produce induced pluripotent stem cells.," *Stem cells (Dayton, Ohio)*, 29(3), pp. 505–16. doi:10.1002/stem.598.

Watt, A.J. *et al.* (2007) "Development of the mammalian liver and ventral pancreas is dependent on GATA4.," *BMC developmental biology*, 7(1), p. 37. doi:10.1186/1471-213X-7-37.

Watt, A.J., Garrison, W.D. and Duncan, S.A. (2003) "HNF4: a central regulator of hepatocyte differentiation and function.," *Hepatology (Baltimore, Md.)*, 37(6), pp. 1249–53. doi:10.1053/jhep.2003.50273.

Weimann, J.M. *et al.* (2003) "Stable reprogrammed heterokaryons form spontaneously in Purkinje neurons after bone marrow transplant," *Nature Cell Biology 2003 5:11*, 5(11), pp. 959–966. doi:10.1038/ncb1053.

Weintraub, Harold *et al.* (1989) "Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD (muscle regulatory gene/MyoD retrovirus)," *Developmental Biology*, 86(14), pp. 5434–5438. doi:10.1073/pnas.86.14.5434.

Weintraub, H. (1993) "The MyoD family and myogenesis: Redundancy, networks, and thresholds," *Cell*, 75(7), pp. 1241–1244. doi:10.1016/0092-8674(93)90610-3.

Weismann, A. (1893) *The Germ-Plasm: A Theory of Heredity*. New York: Charles Scribner's Sons.

Xia, B. and Yanai, I. (2019) "A periodic table of cell types," *Development (Cambridge)*, 146(12). doi:10.1242/DEV.169854/19451.

Xiang, Y. *et al.* (2017) "Fusion of Regionally Specified hPSC-Derived Organoids Models Human Brain Development and Interneuron Migration," *Cell Stem Cell*, 21(3), pp. 383-398.e7. doi:10.1016/J.STEM.2017.07.007.

Xie, H. *et al.* (2004) "Stepwise Reprogramming of B Cells into Macrophages," *Cell*, 117(5), pp. 663–676. doi:10.1016/S0092-8674(04)00419-2.

Xing, Q.R. *et al.* (2020) "Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing.," *Science advances*, 6(37), p. 18. doi:10.1126/sciadv.aba1190.

Xu, J., Du, Y. and Deng, H. (2015) "Direct Lineage Reprogramming: Strategies, Mechanisms, and Applications," *Cell Stem Cell*, 16(2), pp. 119–134. doi:10.1016/J.STEM.2015.01.013.

Yang, J. *et al.* (2019) "Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming," *Stem Cell Reports*, 12(4), pp. 757–771. doi:10.1016/J.STEMCR.2019.02.010.

Yu, J. *et al.* (2007) "Induced pluripotent stem cell lines derived from human somatic cells.," *Science (New York, N.Y.)*, 318(5858), pp. 1917–20. doi:10.1126/science.1151526.

Zaret, K.S. (2020) "Pioneer Transcription Factors Initiating Gene Network Changes," *Annual Review of Genetics*, 54, pp. 367–385. doi:10.1146/ANNUREV-GENET-030220-015007.

Zaret, K.S. and Mango, S.E. (2016) "Pioneer transcription factors, chromatin dynamics, and cell fate control," *Current Opinion in Genetics & Development*, 37, pp. 76–81. doi:10.1016/J.GDE.2015.12.003.

Zeisel, A. *et al.* (2015) "Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.," *Science (New York, N.Y.)*, 347(6226), pp. 1138–42. doi:10.1126/science.aaa1934.

Zeisel, A. *et al.* (2018) "Molecular Architecture of the Mouse Nervous System," *Cell*, 174(4), pp. 999-1014.e22. doi:10.1016/J.CELL.2018.06.021.

Zemmour, D. *et al.* (2018) "Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR," *Nature Immunology 2018 19:3*, 19(3), pp. 291–301. doi:10.1038/s41590-018-0051-0.

Zernicka-Goetz, M., Morris, S.A. and Bruce, A.W. (2009) "Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo," *Nature Reviews Genetics*, 10(7), pp. 467–477. doi:10.1038/nrg2564.

Zhang, Q.-J. *et al.* (2017) "Modeling the phenotype of spinal muscular atrophy by the direct conversion of human fibroblasts to motor neurons.," *Oncotarget*, 8(7), pp. 10945–10953. doi:10.18632/oncotarget.14641.

Zhang, W., Behringer, R.R. and Olson, E.N. (1995) "Inactivation of the myogenic bHLH gene MRF4 results in up-regulation of myogenin and rib anomalies.," *Genes & Development*, 9(11), pp. 1388–1399. doi:10.1101/GAD.9.11.1388.

Zhang, Y. *et al.* (2014) "An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex," *Journal of Neuroscience*, 34(36), pp. 11929–11947. doi:10.1523/JNEUROSCI.1860-14.2014.

Zhao, R. *et al.* (2005) "GATA6 is essential for embryonic development of the liver but dispensable for early heart formation.," *Molecular and cellular biology*, 25(7), pp. 2622–31. doi:10.1128/MCB.25.7.2622-2631.2005.

Zhou, Q. *et al.* (2008) "In vivo reprogramming of adult pancreatic exocrine cells to beta-cells," *Nature*, 455(7213), pp. 627–632. doi:10.1038/NATURE07314.

Zhu, C., Preissl, S. and Ren, B. (2020) "Single-cell multimodal omics: the power of many," *Nature Methods 2020 17:1*, 17(1), pp. 11–14. doi:10.1038/s41592-019-0691-5.

Zunder, E.R. *et al.* (2015) "A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry," *Cell Stem Cell*, 16(3), pp. 323–337. doi:10.1016/j.stem.2015.01.015.

LIST OF PUBLICATIONS

Pircs, K., Petri, R., Madsen, S., Brattås, P.L., Vuono, R., Ottoson, D.R., St-Amour, I., **Hersbach, B.A.,** Matusiak-Brückner, M., Hult Lundh, S., Petersén, A., Déglon, N., Hébert, S.S., Parmar, M., Barker, R.A & Jakobsson, J. (2018). Huntingtin Aggregation Impairs Autophagy, Leading to Argonaute-2 Accumulation and Global MicroRNA Dysregulation. *Cell Reports,* 24(6), 1397-1406.

Brattås, P.L., **Hersbach, B.A.,** Madsen, S., Petri, R., Jakobsson, J., & Pircs, K. (2020). Impact of differential and time-dependent autophagy activation on therapeutic efficacy in a model of Huntington disease. *Autophagy*, 17(6), 1316-1329.

Pircs, K., Drouin-Ouellet, J., Horváth, V., Gil, J., Rezeli, M., Garza, R., Grassi, D.A., Sharma, Y., St-Armour, I., Harris, K., Jönsson, M.E., Johansson, P.A., Vuono, R., Fazal, S.V., Stoker, T., **Hersbach, B.A.,** Sharma, K., Lagerwall, J., Lagerström, S., Storm, P., Hébert, S., Marko-Varga, G., Parmar, M, Barker, R.A., & Jakobsson, J. (2021). *Brain*, awab473.

Kempf, J., Knelles, K., **Hersbach, B.A.,** Petrik, D., Riedemann, T., Bednarova, V., Janjic, A., Simon-Ebert, T., Enard, W., Smialowski, P., Götz, M., & Masserdotti, G. (2021). Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneuronal factors Ascl1 and Neurogenin2. *Cell Reports,* 36(3), 109409.

Hersbach, B.A., Fischer, D.S., Masserdoti, G., Deeksha., Mojžišová, K., Waltzhöni, T, Rodrigues-Terrones, D., Heinig, M., Theis, F.J., Götz, M., Stricker, S.H. (2022). Probing cell identity hierarchies by fate titration and collision during direct reprogramming. *Manuscript submitted to Nature Cell Biology on 08.02.2022.*

EIDESSTATTLICHE VERSICHERUNG/AFFIDAVIT

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation "**Resolving reprogramming factor and starter cell impact on cell fate conversion**" selbstständig angefertigt habe, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

I hereby confirm that the dissertation "**Resolving reprogramming factor and starter cell impact on cell fate conversion**" is the result of my own work and that I have only used sources or materials listed and specified in the dissertation.

München/Munich, 01.03.2022 Bobby Aron Hersbach

DECLARATION OF AUTHOR CONTRIBUTIONS

Kempf, J., Knelles, K., **Hersbach, B.A.,** Petrik, D., Riedemann, T., Bednarova, V., Janjic, A., Simon-Ebert, T., Enard, W., Smialowski, P., Götz, M., & Masserdotti, G. (2021). Heterogeneity of neurons reprogrammed from spinal cord astrocytes by the proneuronal factors Ascl1 and Neurogenin2. *Cell Reports,* 36(3), 109409.

G.M. conceived and designed the project. J.K. and K.K. performed the isolation, culturing, and most of the reprogramming experiments. J.K. and K.K. evaluated the reprogramming efficiency at 8DPI; J.K. performed and quantified the experiments at day 21. K.K. prepared the cells for FACS analysis and extracted RNA for bulk RNA-seq analysis. **B.A.H.** prepared primary cultures of astrocytes from cortex GM, collected single cells, processed the samples with SmartSeq2, and prepared the libraries for sequencing; D.P. collected single iNs for patch-seq, processed them via Smart-Seq2 protocol, and performed morphometric analysis. V.B performed patch-seq experiments, analyzed the electrophysiological data, prepared the samples via Smart-Seq2, and performed initial analysis of RNA-seq data. A.J. prepared bulk-adapted mcSCRB-seq libraries; T.S.-E. contributed to establish MACS protocol for SC, trained J.K. and K.K., and performed immunocytochemistry; T.R. performed whole-cell patch-clamp experiments and analyzed the electrophysiological properties of iNs. W.E. provided reagents; G.M. performed FACS and analyzed data from bulk and single-cell RNA-seq; P.S. compared single-cell RNA with publicly available data and analyzed single-cell RNA-seq with G.M.; M.G. contributed to the conceptual organization of the manuscript and the patch-seq data and financed the work. G.M. and M.G. wrote the manuscript, and **all authors contributed corrections and comments.**

Hersbach, B.A., Fischer, D.S., Masserdoti, G., Deeksha., Mojžišová, K., Waltzhöni, T, Rodrigues-Terrones, D., Heinig, M., Theis, F.J., Götz, M., Stricker, S.H. (2022). Probing cell identity hierarchies by fate titration and collision during direct reprogramming. *Manuscript submitted to Nature Cell Biology on 08.02.2022.*

S.H.S conceived this project. The experimental approach was designed by S.H.S and **B.A.H** together with M.G., F.J.T., G.M. and D.S.F. **B.A.H.** performed all cloning and biological experiments. T.W., D.R. and M.H. performed genome alignment for sequencing data. D.S.F. and K.M performed all bioinformatic analysis. S.H.S, D.S.F and **B.A.H.** wrote the manuscript and all authors contributed to corrections and comments.

Johannes Kempf	Klara Knelles	Bobby Aron Hersbach (Student)
David S. Fischer		Prof.Dr. Magdalena Götz (Supervisor)