HYPOTHESIS

Neuronal differentiation strategies: insights from single-cell sequencing and machine learning

Nikolaos Konstantinides* and Claude Desplan*

ABSTRACT

Neuronal replacement therapies rely on the *in vitro* differentiation of specific cell types from embryonic or induced pluripotent stem cells, or on the direct reprogramming of differentiated adult cells via the expression of transcription factors or signaling molecules. The factors used to induce differentiation or reprogramming are often identified by informed guesses based on differential gene expression or known roles for these factors during development. Moreover, differentiation protocols usually result in partly differentiated cells or the production of a mix of cell types. In this Hypothesis article, we suggest that, to overcome these inefficiencies and improve neuronal differentiation protocols, we need to take into account the developmental history of the desired cell types. Specifically, we present a strategy that uses single-cell sequencing techniques combined with machine learning as a principled method to select a sequence of programming factors that are important not only in adult neurons but also during differentiation.

KEY WORDS: *In vitro* differentiation, Machine learning, Neuronal development, Neuronal differentiation protocols, Neuronal replacement therapy, Single-cell sequencing

Introduction

The goal of neuronal replacement therapy is to provide differentiated neurons to the brain. These are either the product of *in vitro* directed differentiation of embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs), or result from the reprogramming (transdifferentiation) of differentiated adult cells into specific neuronal cell types. Establishing these protocols relies on the expression of a few transcription factors or signaling molecules (termed programming or reprogramming factors), or on treatment with cytokines, growth factors or small molecules (Kidder, 2014; Pfisterer et al., 2016; Huch et al., 2017). The results are variable; not all cells acquire the desired fate and even those that do vary in the extent of their differentiation. This has direct negative consequences on any potential neuronal replacement therapy.

Understanding how transcription factors and signaling molecules are utilized in diverse cell types to orchestrate their specification and differentiation is fundamental to achieve robust and efficient protocols for regenerative medicine. Although the idea of defining *in vitro* differentiation protocols that follow similar steps to embryonic development is not new (Cohen and Melton, 2011), we suggest here that, in light of recent advances in developmental neurobiology, new approaches can be employed to enhance *in vitro* differentiation and transdifferentiation efficiency. We discuss the

Department of Biology, New York University, New York, NY 10003, USA.

*Authors for correspondence (nk1845@nyu.edu; cd38@nyu.edu)

N.K., 0000-0002-9525-2838; C.D., 0000-0002-6914-1413

concept of 'phenotypic convergence' and the genetic and epigenetic mechanisms that explain it, and we highlight how it is one of the reasons for pitfalls in current approaches. We argue that it would be beneficial to change the design of differentiation protocols, using single-cell RNA sequencing approaches and machine-learning algorithms, to transition from informed guesses of programming and reprogramming factors to algorithmically chosen methods.

Current neuronal differentiation strategies

Neurons are characterized by a number of different features that support their function: they express specific adhesion molecules that ensure synaptic specificity; they acquire specific morphologies and target different regions of the body; they form chemical or electric synapses; they use a range of different neurotransmitters to deliver a signal and express different neurotransmitter receptors to receive and propagate signals; and they secrete different signaling proteins. To acquire these features faithfully, neurons express transcription factors that regulate their structural, molecular and physiological characteristics and generate their impressive cell type diversity. These transcription factors control different features at different times during development and are thus expressed upon neuronal specification and/or during differentiation (Fig. 1). Indeed, the nature of these factors and the order in which they are expressed are fundamental for a neuron to acquire its features.

Current neuronal differentiation protocols rely on the synchronous or serial supply of a handful of transcription factors and signaling molecules to ESCs or iPSCs (Fig. 1). Alternatively, already differentiated cells (often non-neuronal) can be directly reprogrammed into a neuronal cell type by expressing one or a few reprogramming factors (Gascón et al., 2017). This conversion of one cell type to another via the misexpression of specific transcription factors was first performed in the 1980s (Davis et al., 1987). However, research in programming and reprogramming cells towards specific cell fates took off after the discovery of the Yamanaka factors, Oct4 (Pou5f1), Sox2, Klf4 and Myc, which were able to reprogram mouse or human fibroblasts into iPSCs. To discover these factors, Takahashi and Yamanaka (2006) screened different combinations of 24 selected factors known to be important for ESC fate and tested them for their efficiency to reprogram fibroblasts into iPSCs. This required a tremendous amount of work, which was rewarded by the 2012 Nobel Prize in Physiology or Medicine. Since then, multiple protocols have been established to reprogram differentiated cells into iPSCs (Takahashi and Yamanaka, 2016), and to differentiate iPSCs or ESCs into specific cell types.

The development of these protocols usually involves two steps: (1) identifying candidate genes that could have a decisive role in the development of the desired cell type; and (2) testing these candidates in different combinations for their efficiency in generating the desired cell type. Together, these protocols have allowed the generation of a variety of neuronal cell types using
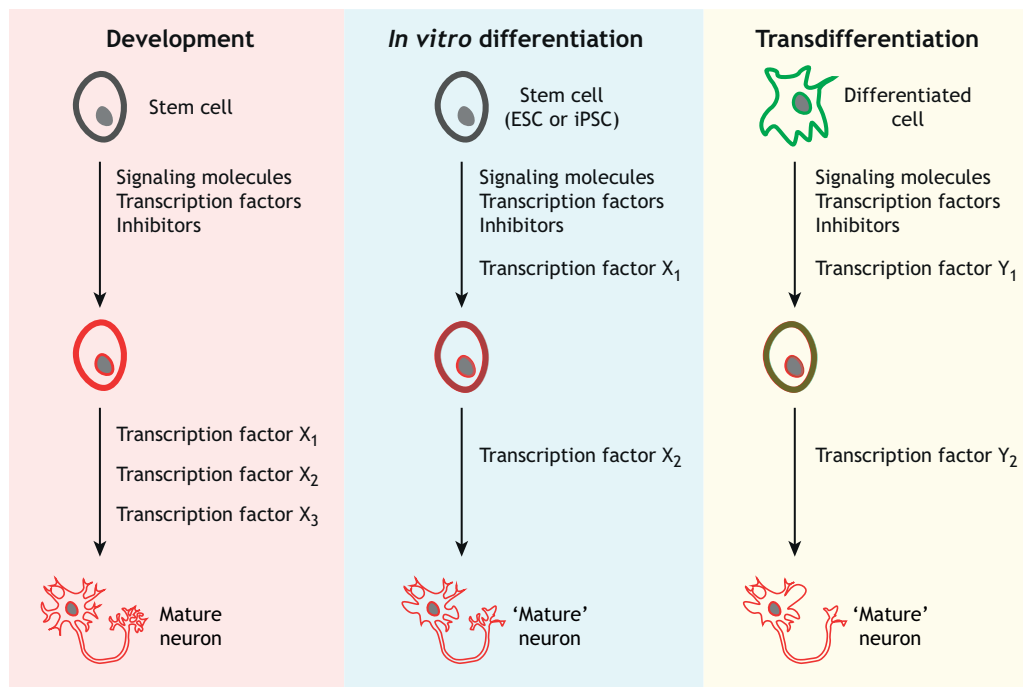
1

Fig. 1. Neuronal differentiation *in vivo* and *in vitro*. During development (left), neuronal stem cells respond to extrinsic cues and express distinct transcription factor combinations in a precise temporal order to generate specific neurons. In an effort to recapitulate this process, *in vitro* differentiation (center) and transdifferentiation (right) protocols supply embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) and differentiated cells, respectively, with a cocktail of signaling molecules and transcription factors that lead to the formation of a 'mature' neuron. These 'mature' neurons are in many respects reminiscent of their wild-type counterparts, but they usually differ in many features, such as maturity, and even neuronal identity. The main reason behind these differences is that the precise temporal pattern of differentiation during embryonic development remains unknown for most neuronal types. Moreover, in the case of transdifferentiation, the cell may progress through intermediate states that could affect the 'mature' phenotype. Finally, the factors used in *in vitro* differentiation or transdifferentiation come mostly from informed guesses.

various starting points, including iPSCs (Hester et al., 2011; Ho et al., 2016; Nehme et al., 2018), ESCs (Hester et al., 2011), neuronal progenitors, fibroblasts (Pfisterer et al., 2016; Son et al., 2011; Wapinski et al., 2013; Xu et al., 2016; Xiao et al., 2018), astrocytes (Corti et al., 2012) or even T cells (Haag et al., 2018), and using a multitude of different neurogenic transcription factors [such as Neurog2, Ascl1, Lmx1a, Brn2 (Pou3f2) etc.]. Indeed, this approach has been applied successfully to induce the production of motor neurons, excitatory cortex neurons, GABAergic neurons, dopaminergic neurons and serotonergic neurons (Caiazzo et al., 2011; Hester et al., 2011; Ho et al., 2016; Xu et al., 2016).

Drawbacks and challenges in neuronal differentiation

Despite huge advances over the last 10 years in generating various neuronal cell types, most of the factors used for directed differentiation have been identified by informed guesses and extensive 'trial and error' approaches. Moreover, they often exhibit low (re)programming efficiency. Notably, three main drawbacks have emerged. First, the efficiency of these protocols, i.e. the percentage of cells that acquire the desired identity, varies from low (<10%; Son et al., 2011) to moderate (~60%; Hester et al., 2011) and in very rare cases is high enough for clinical utilization. Second, even if the efficiency of programming were to be 100%, the cells generated do not belong to one specific cell type but rather represent a broad collection of related cell types that are often not completely differentiated. Recent single-cell analyses have uncovered the amazing neuronal type diversity of the human brain. Therefore, genetic protocols that are able to differentiate ESCs and iPSCs into cholinergic, GABAergic or glutamatergic

neurons are destined to generate a variety of these cell types. Finally, *in vitro* generated neurons often do not correspond accurately to any cell type found to occur naturally within a primary tissue (La Manno et al., 2016). The value of these protocols for *in vivo* programming or reprogramming for neuronal replacement in the clinic therefore remains questionable. The only approaches close to clinical application are those that involve *in vitro* differentiation followed by purification and transplantation, such as that of fetal progenitor cells transformed into dopaminergic neurons to treat Parkinson's disease (Kefalopoulou et al., 2014; Barker et al., 2013) or of ESCs to retinal pigment epithelium to treat age-related macular degeneration (Schwartz et al., 2012; Schwartz et al., 2015). Nonetheless, these protocols have proven to be useful for modeling diseases *in vitro* and for studying these cell types in culture where the successfully programmed cells can be selected based on marker gene expression.

Moving forward, these and other challenges need to be overcome in order to achieve the goal of neuronal replacement therapy. As the brain is composed of a huge number of different cell types, each of which exhibits a unique developmental history and function, it is necessary to know how exact neuronal cell types are specified during development in order to recapitulate their development *in vitro*. In addition, how the efficiency of a given protocol is evaluated currently varies and needs to be addressed. Programming efficiency has generally been measured by assessing the expression of generic neuronal markers, such as Tuj1 (Tubb3), MAP2, NeuN (Rbfox3), Syt1, Syn1, as well as specific markers for serotonergic, dopaminergic or other neuronal types. But how the levels of these translate into neuronal identity and function *in vivo* remains unclear. Moreover, the functionality of a differentiated neuron is often

measured by its capacity to produce spike trains or form synapses but, again, this does not necessarily mean that it is capable of performing the exact function of its natural counterpart.

Phenotypic convergence and *in vitro* differentiation

It is striking that multiple very distinct protocols are capable of programming stem cells or converting differentiated cells into a given cell type. After the four Yamanaka factors were discovered, a number of papers identified other combinations of pluripotency factors that are capable of inducing reprogramming, including members of the Sox, Klf and Myc families, along with Nanog, Lin28, Glis1 and others (Yu et al., 2007; Nakagawa et al., 2008; Abdelalim et al., 2014; Bourillot and Savatier, 2010; Maekawa et al., 2011). The same is true for factors that promote the *in vitro* differentiation of ESCs and iPSCs towards hepatocytes, cardiomyocytes and neurons. In fact, a recent study (Tsunemoto et al., 2018) evaluated 598 pairs of transcription factors (from a pool of 59 transcription factors) for their capacity to reprogram fibroblasts into neurons; 76 of the tested pairs (i.e. 13%) were able to reprogram fibroblasts into neuronal cells that expressed the expected neuronal markers, exhibited neuronal morphology, produced spike trains, and formed synapses. This raises very important questions: does this mean that there are many ways to generate functional neurons? What types of neurons are produced and are they all equivalent? Are they all candidates for neuronal replacement therapies?

This finding also illustrates the notion that the generation of specific neuronal features in a given cell type can be achieved by different regulatory mechanisms and through different routes of differentiation. This phenomenon is called phenotypic convergence (Fig. 2): two cell types can achieve the same phenotype (e.g. morphological, physiological or molecular characteristics) following different developmental paths and using different regulators (Konstantinides et al., 2018). The phenomenon of phenotypic convergence is observed in several *in vivo* contexts. For example, cholinergic neurons in *Caenorhabditis elegans* can be generated *in vivo* in different ways, i.e. different transcription factors are employed in different neuronal types to drive the expression of the cholinergic gene battery. The same is true for *C. elegans* GABAergic (Gendrel et al., 2016) and glutamatergic gene batteries. *Drosophila* optic lobes also employ different transcription factors to generate neurons with the same neurotransmitters (Fig. 2), and to drive the expression of other broadly expressed genes (Konstantinides et al., 2018). In extreme cases in *C. elegans*, the exact same cell type can be generated through distinct developmental paths (Mizeracka et al., 2019 preprint). Phenotypic convergence is also likely to exist in vertebrate neurons, which would explain why distinct differentiation cocktails can be used to generate similarly looking neuronal types. But the question remains: are the neurons generated by different cocktails the same cell type? Furthermore, although they might exhibit shared expression of particular markers, how much do these cell types differ?

The genetic and epigenetic landscape of neurons with convergent characters

Although two cell types may converge on a number of specific characters (e.g. having the same neurotransmitter), it should be noted that these two cell types are by no means equivalent. They will have followed alternative routes during their differentiation and this will undoubtedly have had consequences on their regulatory landscapes. Although these consequences may not be immediately clear from the specific markers that are tested at the end of the differentiation protocol, it is essential to understand them if these cells are to be used as a therapeutic resource.

When two differentiation cocktails are used to direct iPSCs towards neuronal identity (Fig. 2), the transcription factors that are supplied in the two cocktails are different, but lead to the expression of generic neuronal markers that direct the cell to acquire neuronal features and activate the expression of the machinery needed to generate action potentials. However, the two cocktails also activate a number of non-overlapping genes that drive the differentiated neurons to differ from each other. For example, although two

Example		Similarities	Differences
C. elegans			Interneuron versus motor neuron
	TTX-3 →	Cholinergic neurons	Different neuropeptide and receptor expression
	UNC-3 →		Other molecular differences
D. melanogaster optic lobes			Different neuronal types
	Traffic jam →	Glutamatergic neurons	Different CAM and neurotransmitter receptor expression
	Fd59a →		Other molecular differences
In vitro neuronal differentiation			Different neuronal fates
	Ascl2	Neuronal identity	Distinct chromatin accessibility and enhancer activity
	Neurog2		Molecular differences driven by the same downstream transcription factors

Fig. 2. Phenotypic convergence. Neurons can use distinct developmental pathways to achieve identical phenotypic features; this phenomenon is referred to as phenotypic convergence. Three examples are shown here. In *C. elegans*, two different transcription factors, encoded by *ttx-3* and *unc-3*, are responsible for generating cholinergic identity in different neuronal types. However, although these neuronal types share neurotransmitter identity, they are different in many other respects: one of them is an interneuron, whereas the other is a motor neuron, and they express different neuropeptides and neuropeptide receptors. Similarly, in the *Drosophila* optic lobes, glutamatergic identity is regulated by Traffic jam and Fd59a in different neuronal types. However, the same transcription factors differentially regulate other aspects of neuronal identity, such as neurotransmitter and cell adhesion molecule (CAM) expression. Finally, both Ascl2 and Neurog2 can confer neuronal identity *in vitro* in embryonic stem cells, but the resulting neurons are very different in terms of their chromatin state, enhancer activity, and neuronal effector expression. These cases of phenotypic convergence highlight the importance of the developmental pathway a neuron takes for its final identity and emphasize that different programming factors may have very different outcomes, despite seemingly leading to the same neuronal type.

different cocktails may both give rise to cholinergic neurons, the neurons that are generated from each protocol may have differences in the cell adhesion molecules or neurotransmitter receptors that they express (Fig. 2). Therefore, depending on which cell type one wants to generate *in vitro*, different transcription factor cocktails must be used.

The different neurons generated by the 76 pairs of transcription factors described above (Tsunemoto et al., 2018) share similarities with regard to a number of neuronal characteristics. Interestingly, several different transcription factor combinations are able to lead to the same neurotransmitter phenotype, providing yet another example of phenotypic convergence. However, the overall transcriptomes of these cells differ significantly (Tsunemoto et al., 2018), as they include different gene regulatory modules, indicating that they correspond to different cell types. Furthermore, when comparing these neurons to endogenous neuronal populations, only limited similarities to any known specific neurons are found. This highlights both the extent of phenotypic convergence, as well as how difficult it is to use candidate transcription factor approaches to identify *in vitro* differentiation cocktails that can generate specific neuronal types precisely.

Aside from their effects on transcriptomes, transcription factor cocktails have pronounced effects on chromatin marks in differentiating neurons. For example, neuronal progenitors in different parts of the mouse central nervous system have different capacities for producing different neuronal types. This appears to be mediated by the differential chromatin accessibility of their genomes (Metzis et al., 2018), which is determined early on in development depending on the spatial location of their progenitors. The same is true for *Drosophila* ventral nerve cord neural stem cells, in which spatial genes that are expressed early establish neuroblast-specific chromatin landscapes for the later-acting temporal transcription factors (Sen et al., 2019). It comes as no surprise that different transcription factors that are used to generate neurons (e.g. Ascl2 versus Neurog2) have very distinct effects on the chromatin landscapes of the generated neurons, which in turn affects the neuronal subtypes produced in protocols that use one gene or the other. For instance, Brn2, Ebf2 and Onecut2 bind to different genomic sites that are highly dependent on the chromatin landscape set up by Ascl2 or by Neurog2 (Aydin et al., 2019), leading to the production of different motor neurons.

Therefore, although two cell types may appear similar, they are not necessarily functionally equivalent. The obvious question that arises is how similar should the *in vitro* differentiated neuronal population be to the natural one to be able to complement it functionally in order to achieve clinical relevance? Looking in non-neuronal systems, we can draw some interesting conclusions. Chondrocytes, for instance, represent a strong example of phenotypic convergence. Seemingly identical chondrocytes can have ectodermal (neural crest, e.g. nasal chondrocytes) or mesodermal (lateral plate mesoderm and paraxial mesoderm, e.g. knee cartilage) origins (Taïhi et al., 2019). Interestingly, transplantation of nasal chondrocytes can restore knee cartilage defects (Taïhi et al., 2019). Although chondrocytes represent a fairly simple cell type with only three different subtypes (Ji et al., 2019) the main role of which (independent of its origin) is to secrete cartilage, this gives us hope that even if cells are not completely identical, they may still replace each other after transplantation. Cardiomyocytes, by contrast, appear to be less able to replace each other functionally. During development, these cells are generated by both first and second heart field progenitors (Später et al., 2014). The pluripotent stem cell-derived cardiomyocytes that are typically used for transplantation are a mixture of ventricular, atrial and nodal cardiomyocytes (Kadota and Shiba, 2019), making it difficult to assess the contribution of each to the cardiac subtypes. However, it has been demonstrated that atrial cells cannot replace ventricular cells functionally when transplanted into the left ventricle, as they retain their unique atrial phenotype (e.g. shorter calcium transient duration) (Rubart et al., 2003). It is thus not yet clear how faithfully *in vitro* differentiated cell types should recapitulate their natural counterparts to be able to substitute them, and this is something that needs to be studied in further detail.

Single-cell sequencing and machine learning as a means to predict core regulatory transcription factors

Providing terminal transcription factors to stem cells in order to program them largely ignores the developmental path of neurons, and we believe that this is the main reason behind the low efficiency of many programming protocols. However, the advent of single-cell sequencing techniques (Zheng et al., 2017), the development of trajectory inference (Trapnell et al., 2014; Cannoodt et al., 2016; Haghverdi et al., 2016; Setty et al., 2016; Nowakowski et al., 2017; Qiu et al., 2017; La Manno et al., 2018; Wolf et al., 2018; Saelens et al., 2019), and the use of machine learning algorithms to analyze large datasets now gives us the opportunity to identify (re)programming factors in a more rigorous way and to improve differentiation protocols. Based on these advances, we suggest six steps (Fig. 3) to develop differentiation strategies that are designed intelligently and can be used for neuronal replacement therapies: (1) perform single-cell sequencing at different developmental stages, from early embryogenesis all the way to a fully developed brain, using developing brains if available or organoids representing regions in which the specific neurons of interest are naturally produced; (2) use these data to build trajectories from neural stem cells to fully differentiated neuronal types of interest; (3) using machine learning, identify the transcriptions factors involved in the adult and developmental core regulatory complexes that are necessary for a given cell type to develop and that define this cell type; (4) use these transcription factors to develop new *in vitro* differentiation protocols; (5) evaluate the induced neurons using single-cell RNA sequencing and compare them with their natural counterparts; (6) depending on the results, use the data from the evaluation to adapt the protocol by improving the machine learning algorithms and repeating steps 4-5.

Step 1: Single-cell sequencing throughout development

A number of different single-cell sequencing platforms, such as 10x Genomics (Zheng et al., 2017) and Smart-Seq (Ramsköld et al., 2012), can be used to generate transcriptomic information from almost every neuronal type, even rare ones in the adult brain. The Human Cell Atlas (Regev et al., 2017) has already set out to generate such data for every adult tissue, and the Allen Institute for Brain Science (Tasic et al., 2018) and the Brain Initiative Cell Census consortium (Ecker et al., 2017) have shown that this is feasible in a short time frame. However, it is clear that many neuronal features, such as the acquisition of appropriate morphology, synaptic partner selection and synapse formation, are established early during development; thus, sequencing adult neuronal types is not sufficient to capture these events. Obviously, obtaining such data from different developmental stages requires a coordinated effort, equivalent to that of the Human Cell Atlas, with the goal of mapping all developmental lineage decisions. A main obstacle to gaining transcriptomic information during development is acquiring access to developing neurons *in situ*. Access to fetal brain tissue
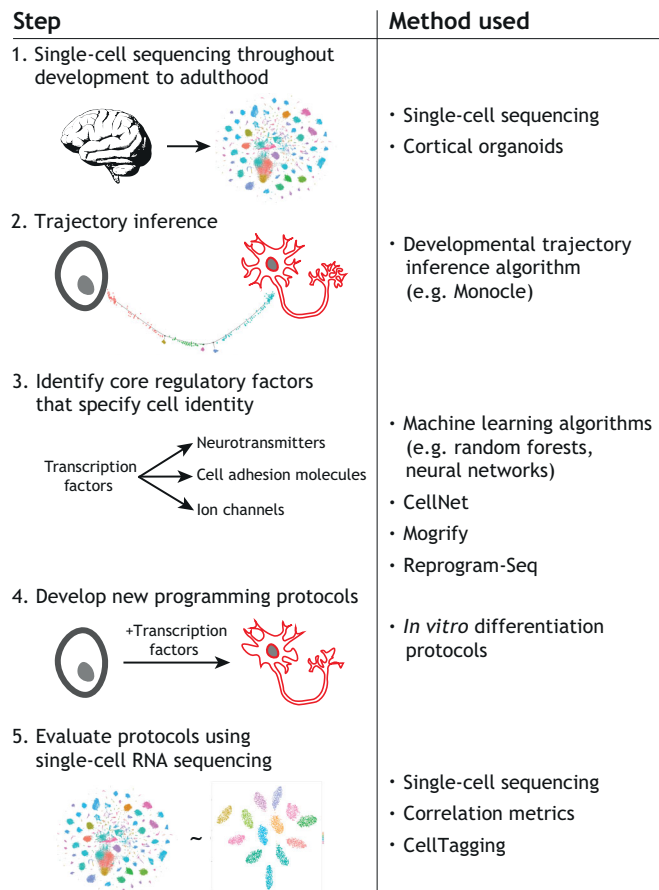
Step	Method used
1. Single-cell sequencing throughout development to adulthood	• Single-cell sequencing • Cortical organoids
2. Trajectory inference	• Developmental trajectory inference algorithm (e.g. Monocle)
3. Identify core regulatory factors that specify cell identity	• Machine learning algorithms (e.g. random forests, neural networks) • CellNet • Mogrify • Reprogram-Seq
4. Develop new programming protocols	• *In vitro* differentiation protocols
5. Evaluate protocols using single-cell RNA sequencing	• Single-cell sequencing • Correlation metrics • CellTagging

Fig. 3. Intelligently designed *in vitro* neuronal differentiation protocols. We propose a protocol that involves six steps. (1) Initially, single-cell sequencing of different neuronal types throughout development has to be performed to provide access to the transcriptomes of the developing neurons. Ideally, the sequencing will be carried out in single cells coming from fetal brain tissue of different ages. Alternatively, one could rely on cortical organoids. (2) Trajectory inference algorithms, such as Monocle, can then be used to build trajectories from neuronal progenitors to fully differentiated neuronal types. (3) Machine learning algorithms, such as those implemented in CellNet, Mogrify and Reprogram-Seq, can then be used to identify the transcription factors that constitute each cell type's developmental and adult core regulatory complex. (4) These transcription factors can then be used to develop new *in vitro* differentiation protocols. (5) These protocols can then be evaluated using single-cell RNA sequencing and by comparing the resultant cells with their natural counterparts. (6) Steps 4 and 5 can be iterated depending on the evaluation to improve the protocols (not shown).

coming from second trimester pregnancy terminations is possible (Nowakowski et al., 2018; Mayer et al., 2019) but is unpredictable as well as being legally and ethically regulated. A potential solution to this involves the use of organoids as a surrogate for human brains. Human brain organoids are three-dimensional structures derived from ESCs or iPSCs. They self-organize and resemble a simplified human brain (Huch et al., 2017). Their main advantage is that they can be used to study human organ development and to recapitulate human diseases, while at the same time being amenable to genetic manipulation; they also hold promise for generating patient-specific models. However, brain organoid technology (Hattori, 2014; Qian et al., 2016; Di Lullo and Kriegstein, 2017; Paşca, 2018; Pollen et al., 2019) is still in its infancy and faces a number of limitations, mainly because organoids cannot recapitulate the sheer complexity of the brain; they do not contain all of the different cell types, they vary from one organoid to the next, and they mature slowly

(Di Lullo and Kriegstein, 2017). Moreover, it is not clear how well organoids recapitulate normal development (Pollen et al., 2019). Nonetheless, recent protocols to generate brain organoids are highly reproducible (Velasco et al., 2019) and can recapitulate the main aspects of cortical development (Pollen et al., 2019), allowing the production of radial glial cells and neurons of all six different layers in a temporal fashion (Kadoshima et al., 2013; Hattori, 2014; Paşca et al., 2015; Qian et al., 2016; Di Lullo and Kriegstein, 2017; Huch et al., 2017). Human brain organoids even form an outer subventricular zone, a region that is missing in mice (Di Lullo and Kriegstein, 2017). The protocols for generating brain organoids are rapidly improving and hopefully with time will provide access to neuronal cell types at different stages of development in order to assemble their developmental trajectories.

Step 2: Trajectory inference
Once single-cell sequencing has been performed, trajectory inference algorithms (Trapnell et al., 2014; Cannoodt et al., 2016; Haghverdi et al., 2016; Setty et al., 2016; Nowakowski et al., 2017; Qiu et al., 2017; La Manno et al., 2018; Saelens et al., 2019; Wolf et al., 2018) allow the ordering of cell states during differentiation processes based on transcriptomic data. It is thus possible to order single cells that belong to the same cell type according to their age or level of maturity. These algorithms can then, for example, define molecules that are differentially expressed during differentiation, cluster genes according to similar expression trends, identify new regulatory dynamics, and pinpoint key differential events. A number of trajectory inference algorithms have been published over the last 5 years (Cannoodt et al., 2016) that allow the discovery of transcription factors (or other key molecules) that are differentially expressed at various stages of neuronal differentiation. The various trajectory inference algorithms available differ in their performance, each of them offering distinct advantages in terms of accuracy and stability of prediction, as well as usability and scalability (Saelens et al., 2019). Although there is no single algorithm that would work well in every dataset, methods such as PAGA (Wolf et al., 2019) and Slingshot (Street et al., 2018) appear to be accurate, reproducible and scalable to large datasets. Using these algorithms, developmental trajectories of cell types of interest in a human brain can be obtained.

Step 3: Identification of core regulatory factors that specify cell type identity
Each cell type is characterized by the expression of a number of transcription factors that can be used to distinguish it from other cell types. Together, these transcription factors generate the cell type-specific characteristics that are necessary for a given cell's functions and represent the core regulatory complex (CoRC) (Arendt et al., 2016) of each cell type. To generate a specific cell type accurately, it is necessary to identify its CoRC at each developmental stage and define how these transcription factors are expressed during and upon successful differentiation (Fig. 4).

But how is it possible to identify the CoRC of each differentiated cell type? First, the transcription factors should be expressed in developing or adult cells, as they are responsible for the generation of differences in gene expression between cell types, e.g. the gene module that is responsible for the generation and release of a particular neurotransmitter at the synapse; these features will be reflected in the transcriptome of the cells. Second, CoRC transcription factors have to be differentially expressed in some cell types compared with others. Such CoRC transcription factors can be identified from single-cell sequencing data using machine
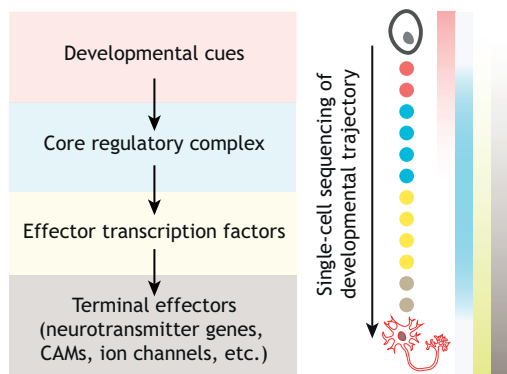
Fig. 4. Identifying the CoRC. The core regulatory complex of transcription factors (CoRC) of a particular cell type interprets developmental cues (in the form of spatiotemporal factors and signaling molecules each cell receives) and establishes cell identity. It then activates the genes that endow each cell with its character: in the case of neurons, these terminal effector genes can be neurotransmitter genes, cell adhesion molecules, receptors, channels, etc. CoRC transcription factors can activate terminal effectors directly or indirectly through effector transcription factors. Molecules at different levels of this differentiation process are expressed with different timings. The temporal progression of differentiation from a specified progenitor to a fully differentiated cell can be studied using single-cell sequencing at different stages of development and reconstructing the developmental trajectory of each cell type.

learning algorithms. These algorithms build mathematical models using training data that allow them to make predictions beyond these data. Different types of mathematical models exist, the simplest being linear regression; other types include decision trees, random forests and support vector machines. These models can be used to infer relationships such as covariance between genes. Covariance between transcription factors and terminal effector genes can be used to infer regulatory interactions between them, which can then be tested genetically. Specifically, one can use the single-cell sequencing data that were generated in step 1 to identify transcription factors and neuronal effector genes that are expressed in different neuronal types. Machine learning algorithms can then identify transcription factors that co-vary with neuronal effector genes and expression of which could predict the identity of the neuronal type (Konstantinides et al., 2018). These would reduce the hundreds of transcription factors that are expressed in each neuronal type to a handful (10-15) that control the acquisition and implementation of neuronal type identity.

A number of such predictive algorithms that were generated to accelerate high-throughput screenings of potential differentiation factors have used expression data from adult differentiated cells to identify cell type-specific differentiation cocktails (Cahan et al., 2014; Rackham et al., 2016; Duan et al., 2019) that perform better than those generated by trial and error. For instance, CellNet (Cahan et al., 2014; Morris et al., 2014) initially relied on 3419 published gene expression profiles of diverse cell types and tissues, such as ESCs, neurons, glia, muscle, fibroblasts, endothelial cells and hematopoietic stem cells, to identify gene regulatory networks that are expressed in specific cell types and are necessary for endowing cells with their correct identity. It was then used to compare the directed differentiation of stem cells to the direct conversion (transdifferentiation) of one cell type to another, and to improve protocols of *in vitro* transdifferentiation of B cells into macrophages (Morris et al., 2014). Mogrify (Rackham et al., 2016) similarly uses gene expression data and gene regulatory information to predict transcription factors required for different cell type conversions (Rackham et al., 2016), allowing the generation of new protocols for

the transdifferentiation of human fibroblasts into keratinocytes and of human keratinocytes into microvascular endothelial cells. Reprogram-Seq (Duan et al., 2019) incorporates the use of single-cell mRNA sequencing and perturbation analysis to more accurately predict and evaluate transcription factor cocktails that can reprogram specific cell types, allowing, for instance, the identification of a new combination of transcription factors that can convert embryonic fibroblasts into epicardial cells.

As a second step, using the developmental trajectories emanating from steps 1 and 2 and machine learning algorithms, we can then identify the crucial transcription factors that are expressed during development to supplement the adult CoRC. The ability of the identified developmental and adult CoRC to define a cell type can then be tested by using these transcription factors to direct the programming of neural stem cells towards specific cell types *in vitro*. This will test both the effectiveness of the CoRC and will be a practical tool.

Steps 4-6: Develop new programming protocols and evaluate using single-cell RNA sequencing

Identifying the developmental and adult CoRC of a specific cell type will allow the development of neuronal differentiation protocols using these transcription factors as programming factors. Although it is not an easy endeavor to co-express numerous factors, a number of techniques, including traditional cDNA overexpression but also more elaborate CRISPR-based multiplexed genome engineering techniques (Campa et al., 2019) that allow for controlled expression or silencing of multiple genes in the same cell, could be used to supply these transcription factors in a synchronous or serial manner. Single-cell sequencing can then be used to evaluate the programming efficiency of any protocol and assess the extent of potential heterogeneities that may emerge during differentiation. It can also be applied at different stages of the *in vitro* differentiation process to construct trajectories of different cell types that can then be compared with the developmental trajectories that occur during embryonic development. More importantly, techniques such as 'CellTagging' (Biddy et al., 2018), which relies on the sequential delivery of heritable barcodes during *in vitro* differentiation, can be used for the simultaneous capture of lineage and cell identity. This can be particularly insightful in cases of lower programming efficiency and can provide an understanding of how undesired cell types arise during the differentiation process. This understanding, in turn, could be used to modify the differentiation protocol to tilt the balance towards the desired cell type, thus increasing the programming efficiency and decreasing nonspecific byproducts.

Other considerations when designing differentiation strategies
Direct reprogramming
Although we have focused here on directed differentiation (i.e. from ESCs or iPSCs), direct reprogramming can also be used to generate specific neuronal cell types (Gascón et al., 2017). This process bypasses the pluripotency step and instead converts one cell type into another through transdifferentiation. However, it is less efficient, because evidence of the initial differentiation of the cell, which is prominently written into its chromatin, is difficult to erase. As such, the cells generated via direct reprogramming are often incompletely converted (i.e. they are hybrid cells) and may remain developmentally immature. The efficiency of direct conversion also, therefore, depends on the starting cell type and its relationship to the target cell type. For instance, direct conversions are not

efficient in crossing germ layer boundaries (Sieweke, 2015; An et al., 2018), although they are not impossible (Vierbuchen et al., 2010; Karow et al., 2012). By contrast, the conversion of astrocytes or other glial cells to neurons is efficient (Berninger et al., 2007; Masserdotti et al., 2016). This highlights that the genetic regulatory context in which the transcription factors are expressed can enhance or repress reprogramming efficiency (Aydin and Mazzoni, 2019); therefore, in an already differentiated cell, the chromatin marks that are specific to the cell of origin must be erased, and the specific transcription factors of the target cell must then be expressed. This dependency of transcription factors on the state of chromatin has been clearly shown in *C. elegans*, where RNAi against a chromatin-regulating factor (*lin-53*) and overexpression of a terminal selector (*che-1*) are together sufficient to allow the conversion of a germ cell into a specific neuronal type (Tursun et al., 2011; Kolundzic et al., 2018). The best option for direct conversion is thus to select a cell type of origin for which the genomic landscape resembles that of the target cell type.

Beyond transcription factors and transcriptomes

It is clear that, although there are constantly expanding ways to overexpress transcription factors or downregulate genes of interest in a spatially and temporally controlled way, it is not so trivial to 'write' and 'erase' chromatin marks that might affect the reprogramming process. However, single-cell chromatin accessibility assays (Buenrostro et al., 2015; Cusanovich et al., 2015, 2018; Cao et al., 2018; Pliner et al., 2018) can now be performed to profile the landscape of different neuronal types and select the best candidate chromatin modifiers. In addition, single-cell epigenomic assays could offer complementary information to the transcriptome. For example, some transcription factors may operate in different ways depending on chromatin accessibility (Velasco et al., 2017). Moreover, some transcription factors, being expressed at low levels, may not be unambiguously recovered with single-cell sequencing, but their effects on chromatin might be more profound. This is in agreement with recent studies showing that single-cell transcriptomes are insufficient to separate progenitor cells that will generate distinct cell lineages (Weinreb et al., 2020). This means that there are some hidden layers of progenitor diversity (e.g. differences in chromatin landscapes and/or transcription factors expressed at low levels) that may affect their capacity to generate different cells and that might not be recovered using single-cell sequencing approaches.

Although the role of transcriptional regulation has been very well studied, post-transcriptional processes also affect cell fate, most notably microRNAs, which participate in different levels of reprogramming (Beh-Pajooh et al., 2018). For example, miR-145 regulates the expression of three of the four Yamanaka factors and is involved in a double-negative loop that controls pluripotency in iPSCs (Xu et al., 2009). SImilarly, a combination of microRNAs – miR-1, -133, -208 and -499a – is able to convert mouse cardiac fibroblasts to functional cardiomyocytes (Jayawardena et al., 2012, 2014). Different single-cell sequencing techniques are available that can detect microRNAs alone (mime-Seq; Alberti et al., 2018) or in conjunction with mRNA (single-cell microRNA-mRNA co-sequencing; Wang et al., 2019). Using these techniques, one can identify miRNAs that may be incorporated into the CoRC of transcription factors to better define and generate a particular neuronal identity.

Conclusions and perspectives

We have presented here a strategy to combine single-cell sequencing with machine learning to identify crucial programming factors that could allow us to recapitulate developmental routes and, ultimately, faithfully program *in vitro* neuronal cell types from stem cells. This strategy, however, only addresses the identification of molecules that are necessary to program a cell efficiently. A number of other considerations should also be taken into account to define improved differentiation protocols. For example, how does one ensure the expression of the necessary differentiation molecules at the appropriate timing and physiological levels? How can one control the environment in which a cell grows (i.e. the cell-cell signaling), which is profoundly different in a dish compared with its tissue of origin? Although this falls beyond the scope of this article, CRISPR-based methods that are constantly being developed can address many of these considerations (Cheng et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013; Qi et al., 2013; Dahlman et al., 2015; Kiani et al., 2015; Boettcher et al., 2018). For example, Cas12a has recently been used for constitutive, conditional, inducible and orthogonal gene editing, whereby dozens of different genes can be independently manipulated (upregulated or downregulated) in a controllable, i.e. conditional and inducible, way (Campa et al., 2019).

Sydney Brenner once said 'I will ask you to mark again that rather typical feature of the development of our subject; how so much progress depends on the interplay of techniques, discoveries and new ideas, probably in that order of decreasing importance' (Brenner, 2002). As we have highlighted here, the development of exciting new techniques over the last few years clearly has the potential to lead to important discoveries and trigger new ideas. The interplay of these will hopefully lead to novel applications that can have a lasting impact and provide means for therapeutic interventions, such as regenerative medicine and neuronal replacement therapy.

References
Abdelalim, E. M., Emara, M. M. and Kolatkar, P. R. (2014). The SOX transcription factors as key players in pluripotent stem cells. *Stem Cells Dev.* **23**, 2687-2699. doi:10.1089/scd.2014.0297

Alberti, C., Manzenreither, R. A., Sowemimo, I., Burkard, T. R., Wang, J., Mahofsky, K., Ameres, S. L. and Cochella, L. (2018). Cell-type specific sequencing of microRNAs from complex animal tissues. *Nat. Methods* **15**, 283-289. doi:10.1038/nmeth.4610

An, N., Xu, H., Gao, W.-Q. and Yang, H. (2018). Direct conversion of somatic cells into induced neurons. *Mol. Neurobiol.* **55**, 642-651. doi:10.1007/s12035-016-0350-0

Arendt, D., Musser, J. M., Baker, C. V. H., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D. et al. (2016). The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744-757. doi:10.1038/nrg.2016.127

Aydin, B. and Mazzoni, E. O. (2019). Cell reprogramming: the many roads to success. *Annu. Rev. Cell Dev. Biol.* **35**, 433-452. doi:10.1146/annurev-cellbio-100818-125127

Aydin, B., Kakumanu, A., Rossillo, M., Moreno-Estellés, M., Garipler, G., Ringstad, N., Flames, N., Mahony, S. and Mazzoni, E. O. (2019). Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nat. Neurosci.* **22**, 897-908. doi:10.1038/s41593-019-0399-y

Barker, R. A., Barrett, J., Mason, S. L. and Björklund, A. (2013). Fetal dopaminergic transplantation trials and the future of neural grafting in Parkinson's disease. *Lancet Neurol.* **12**, 84-91. doi:10.1016/S1474-4422(12)70295-8

DEVELOPMENT

Beh-Pajooh, A., Cantz, T., Koch, U., Schroeder, T., Sutor, B., Grothe, B. and Gotz, M. (2018). The role of microRNAs in embryonic and induced pluripotency. *J. Stem Cells Regener. Med.* **14**, 3-9. doi:10.46582/jsrm.1401002

Berninger, B., Costa, M. R., Koch, U., Schroeder, T., Sutor, B., Grothe, B. and Gotz, M. (2007). Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *J. Neurosci.* **27**, 8654-8664. doi:10.1523/JNEUROSCI.1615-07.2007

Biddy, B. A., Kong, W., Kamimoto, K., Guo, C., Waye, S. E., Sun, T. and Morris, S. A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224. doi:10.1038/s41586-018-0744-4

Boettcher, M., Tian, R., Blau, J. A., Markegard, E., Wagner, R. T., Wu, D., Mo, X., Biton, A., Zaitlen, N., Fu, H. et al. (2018). Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170-178. doi:10.1038/nbt.4062

Bourillot, P.-Y. and Savatier, P. (2010). Krüppel-like transcription factors and control of pluripotency. *BMC Biol.* **8**, 125. doi:10.1186/1741-7007-8-125

Brenner, S. (2002). Life sentences: detective rummage investigates. *Genome Biol.* **3**, comment1013.1-comment1013.2.

Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y. and Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-490. doi:10.1038/nature14590

Cahan, P., Li, H., Morris, S. A., Lummertz da Rocha, E., Daley, G. Q. and Collins, J. J. (2014). CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903-915. doi:10.1016/j.cell.2014.07.020

Caiazzo, M., Dell'Anno, M. T., Dvoretskova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T. D., Menegon, A., Roncaglia, P., Colciago, G. et al. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* **476**, 224-227. doi:10.1038/nature10284

Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. and Platt, R. J. (2019). Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* **16**, 887-893. doi:10.1038/s41592-019-0508-6

Cannoodt, R., Saelens, W. and Saeys, Y. (2016). Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* **46**, 2496-2506. doi:10.1002/eji.201646347

Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., Daza, R. M., Mcfaline-Figueroa, J. L., Packer, J. S., Christiansen, L. et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380-1385. doi:10.1126/science.aau0730

Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., Rangarajan, S., Shivalila, C. S., Dadon, D. B. and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163-1171. doi:10.1038/cr.2013.122

Cohen, D. E. and Melton, D. (2011). Turning straw into gold: directing cell fate for regenerative medicine. *Nat. Rev. Genet.* **12**, 243-252. doi:10.1038/nrg2938

Corti, S., Nizzardo. M., Simone, C., Falcone, M., Donadoni, C., Salani, S., Rizzo, F., Nardini, M., Riboldi, G., Magri, F. et al. (2012). Direct reprogramming of human astrocytes into neural stem cells and neurons. *Exp. Cell Res.* **318**, 1528-1541. doi:10.1016/j.yexcr.2012.02.040

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. and Shendure, J. (2015). Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914. doi:10.1126/science.aab1601

Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., Filippova, G. N., Huang, X., Christiansen, L., Dewitt, W. S. et al. (2018). A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309-1324.e18. doi:10.1016/j.cell.2018.06.052

Dahlman, J. E., Abudayyeh, O. O., Joung, J., Gootenberg, J. S., Zhang, F. and Konermann, S. (2015). Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159-1161. doi:10.1038/nbt.3390

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

Di Lullo, E. and Kriegstein, A. R. (2017). The use of brain organoids to investigate neural development and disease. *Nat. Rev. Neurosci.* **18**, 573-584. doi:10.1038/nrn.2017.107

Duan, J., Li, B., Bhakta, M., Xie, S., Zhou, P., Munshi, N. V. and Hon, G. C. (2019). Rational reprogramming of cellular states by combinatorial perturbation. *Cell Rep.* **27**, 3486-3499. doi: 10.1016/j.celrep.2019.05.079

Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P., Polioudakis, D., Regev, A., Sestan, N., Wickersham, I. R. and Zeng, H. (2017). The BRAIN initiative cell census consortium: lessons learned toward generating a comprehensive brain cell atlas. *Neuron* **96**, 542-557. doi:10.1016/j.neuron.2017.10.007

Gascón, S., Masserdotti, G., Russo, G. L. and Götz, M. (2017). Direct neuronal reprogramming: achievements, hurdles, and new roads to success. *Cell Stem Cell* **21**, 18-34. doi:10.1016/j.stem.2017.06.011

Gendrel, M., Atlas, E. G. and Hobert, O. (2016). A cellular and regulatory map of the GABAergic nervous system of C. elegans. *eLife* **5**, e17686. doi:10.7554/eLife.17686

Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845-848. doi:10.1038/nmeth.3971

Hattori, N. (2014). Cerebral organoids model human brain development and microcephaly. *Mov. Disord.* **29**, 185-185. doi:10.1002/mds.25740

Hester, M. E., Murtha, M. J., Song, S. W., Rao, M., Miranda, C. J., Meyer, K., Tian, J., Boulting, G., Schaffer, D. V., Zhu, M. X. et al. (2011). Rapid and efficient generation of functional motor neurons from human pluripotent stem cells using gene delivered transcription factor codes. *Mol. Ther.* **19**, 1905-1912. doi:10.1038/mt.2011.135

Ho, S.-M., Hartley, B. J., TCW, J., Beaumont, M., Stafford, K., Slesinger, P. A. and Brennand, K. J. (2016). Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* **101**, 113-124. doi:10.1016/j.ymeth.2015.11.019

Huch, M., Knoblich, J. A., Lutolf, M. P. and Martinez-Arias, A. (2017). The hope and the hype of organoid research. *Development* **144**, 938-941. doi:10.1242/dev.150201

Jayawardena, T. M., Egemnazarov, B., Finch, E. A., Zhang, L., Payne, J. A., Pandya, K., Zhang, Z., Rosenberg, P., Mirotsou, M. and Dzau, V. J. (2012). MicroRNA-mediated in vitro and in vivo direct reprogramming of cardiac fibroblasts to cardiomyocytes. *Circ. Res.* **110**, 1465-1473. doi:10.1161/CIRCRESAHA.112.269035

Jayawardena, T., Mirotsou, M. and Dzau, V. J. (2014). Direct reprogramming of cardiac fibroblasts to cardiomyocytes using MicroRNAs. *Methods Mol. Biol.* **1150**, 263-272. doi:10.1007/978-1-4939-0512-6_18

Ji, Q., Zheng, Y., Zhang, G., Hu, Y., Fan, X., Hou, Y., Wen, L., Li, L., Xu, Y., Wang, Y. et al. (2019). Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. *Ann. Rheum. Dis.* **78**, 100-110. doi:10.1136/annrheumdis-2017-212863

Kadoshima, T., Sakaguchi, H., Nakano, T., Soen, M., Ando, S., Eiraku, M. and Sasai, Y. (2013). Self-organization of axial polarity, inside-out layer pattern, and species-specific progenitor dynamics in human ES cell-derived neocortex. *Proc. Natl Acad. Sci. USA* **110**, 20284-20289. doi:10.1073/pnas.1315710110

Kadota, S. and Shiba, Y. (2019). Pluripotent stem cell-derived cardiomyocyte transplantation for heart disease treatment. *Curr. Cardiol. Rep.* **21**, 73. doi:10.1007/s11886-019-1171-3

Karow, M., Sánchez, R., Schichor, C., Masserdotti, G., Ortega, F., Heinrich, C., Gascón, S., Khan, M. A., Lie, D. C., Dellavalle, A. et al. (2012). Reprogramming of pericyte-derived cells of the adult human brain into induced neuronal cells. *Cell Stem Cell* **11**, 471-476. doi:10.1016/j.stem.2012.07.007

Kefalopoulou, Z., Politis, M., Piccini, P., Mencacci, N., Bhatia, K., Jahanshahi, M., Widner, H., Rehncrona, S., Brundin, P., Björklund, A. et al. (2014). Long-term clinical outcome of fetal cell transplantation for Parkinson disease: two case reports. *JAMA Neurol.* **71**, 83-87. doi:10.1001/jamaneurol.2013.4749

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S. et al. (2015). Cas9 gRNA engineering for genome editing, activation and repression. *Nat. Methods* **12**, 1051-1054. doi:10.1038/nmeth.3580

Kidder, B. L. (2014). Derivation and manipulation of trophoblast stem cells from mouse blastocysts. *Methods Mol. Biol.* **1150**, 201-212. doi:10.1007/978-1-4939-0512-6_13

Kolundzic, E., Seelk, S. and Tursun, B. (2018). Application of RNAi and heat-shock-induced transcription factor expression to reprogram germ cells to neurons in C. elegans. *J. Vis. Exp.* **131**, e56889. doi:10.3791/56889

Konstantinides, N., Kapuralin, K., Fadil, C., Barboza, L., Satija, R. and Desplan, C. (2018) Phenotypic convergence: distinct transcription factors regulate common terminal features. *Cell* **174**, 622-635.e13. doi:10.1016/j.cell.2018.05.021

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L. E., Stott, S. R. W., Toledo, E. M., Villaescusa, J. C. et al. (2016). Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566-580.e19. doi:10.1016/j.cell.2016.09.027

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A. et al. (2018). RNA velocity of single cells. *Nature* **560**, 494-498. doi:10.1038/s41586-018-0414-6

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N. and Yamanaka, S. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. *Nature* **474**, 225-229. doi:10.1038/nature10106

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H. and Joung, J. K. (2013). CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977-979. doi:10.1038/nmeth.2598

Masserdotti, G., Gascón, S. and Götz, M. (2016). Direct neuronal reprogramming: learning from and for development. *Development* **143**, 2494-2510. doi:10.1242/dev.092163

Mayer, S., Chen, J., Velmeshev, D., Mayer, A., Eze, U. C., Bhaduri, A., Cunha, C. E., Jung, D., Arjun, A., Li, E. et al. (2019). Multimodal single-cell analysis

reveals physiological maturation in the developing human neocortex. *Neuron* **102**, 143-158.e7. doi:10.1016/j.neuron.2019.01.027

Metzis, V., Steinhauser, S., Pakanavicius, E., Gouti, M., Stamataki, D., Ivanovitch, K., Watson, T., Rayon, T., Mousavy Gharavy, S. N., Lovell-Badge, R. et al. (2018). Nervous system regionalization entails axial allocation before neural differentiation. *Cell* **175**, 1105-1118.e17. doi:10.1016/j.cell.2018.09.040

Mizeracka, K., Rogers, J. M., Shaham, S., Bulyk, M. L. and Heiman, M. G. (2019). Lineage-specific control of convergent cell identity by a Forkhead repressor. *bioRxiv*, 758508. doi:10.1101/758508

Morris, S. A., Cahan, P., Li, H., Zhao, A. M., San Roman, A. K., Shivdasani, R. A., Collins, J. J. and Daley, G. Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell* **158**, 889-902. doi:10.1016/j.cell.2014.07.021

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26**, 101-106. doi:10.1038/nbt1374

Nehme, R., Zuccaro, E., Ghosh, S. D., Li, C., Sherwood, J. L., Pietilainen, O., Barrett, L. E., Limone, F., Worringer, K. A., Kommineni, S. et al. (2018). Combining NGN2 programming with developmental patterning generates human excitatory neurons with NMDAR-mediated synaptic transmission. *Cell Rep.* **23**, 2509-2523. doi:10.1016/j.celrep.2018.04.066

Nowakowski, T. J., Bhaduri, A., Pollen, A. A., Alvarado, B., Mostajo-Radji, M. A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S. J., Velmeshev, D. et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323. doi:10.1126/science.aap8809

Nowakowski, T. J., Rani, N., Golkaram, M., Zhou, H. R., Alvarado, B., Huch, K., West, J. A., Leyrat, A., Pollen, A. A., Kriegstein, A. R. et al. (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat. Neurosci.* **21**, 1784-1792. doi:10.1038/s41593-018-0265-3

Paşca, S. P. (2018). The rise of three-dimensional human brain cultures. *Nature* **553**, 437-445. doi:10.1038/nature25032

Paşca, A. M., Sloan, S. A., Clarke, L. E., Tian, Y., Makinson, C. D., Huber, N., Kim, C. H., Park, J.-Y., O'rourke, N. A., Nguyen, K. D. et al. (2015). Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671-678. doi:10.1038/nmeth.3415

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W. et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973-976. doi:10.1038/nmeth.2600

Pfisterer, U., Ek, F., Lang, S., Soneji, S., Olsson, R. and Parmar, M. (2016). Small molecules increase direct neural conversion of human fibroblasts. *Sci. Rep.* **6**, 38290. doi:10.1038/srep38290

Pliner, H. A., Packer, J. S., McFaline-Figueroa, J. L., Cusanovich, D. A., Daza, R. M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A. et al. (2018). Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858-871.e8. doi:10.1016/j.molcel.2018.06.044

Pollen, A. A., Bhaduri, A., Andrews, M. G., Nowakowski, T. J., Meyerson, O. S., Mostajo-Radji, M. A., Di Lullo, E., Alvarado, B., Bedolli, M., Dougherty, M. L. et al. (2019). Establishing cerebral organoids as models of human-specific brain evolution. *Cell* **176**, 743-756.e17. doi:10.1016/j.cell.2019.01.017

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183. doi:10.1016/j.cell.2013.02.022

Qian, X., Nguyen, H. N., Song, M. M., Hadiono, C., Ogden, S. C., Hammack, C., Yao, B., Hamersky, G. R., Jacob, F., Zhong, C. et al. (2016). Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* **165**, 1238-1254. doi:10.1016/j.cell.2016.04.032

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A. and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979-982. doi:10.1038/nmeth.4402

Rackham, O. J. L., 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. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* **48**, 331-335. doi:10.1038/ng.3487

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O. R., Daniels, G. A., Khrebtukova, I., Loring, J. F., Laurent, L. C. et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777-782. doi:10.1038/nbt.2282

Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M. et al. (2017). The human cell atlas. *eLife* **6**, e27041. doi:10.7554/eLife.27041

Rubart, M., Pasumarthi, K. B. S., Nakajima, H., Soonpaa, M. H., Nakajima, H. O. and Field, L. J. (2003). Physiological coupling of donor and host cardiomyocytes after cellular transplantation. *Circ. Res.* **92**, 1217-1224. doi:10.1161/01.RES.0000075089.39335.8C

Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547-554. doi:10.1038/s41587-019-0071-9

Schwartz, S. D., Hubschman, J. P., Heilwell, G., Franco-Cardenas, V., Pan, C. K., Ostrick, R. M., Mickunas, E., Gay, R., Klimanskaya, I. and Lanza, R. (2012). Embryonic stem cell trials for macular degeneration: a preliminary report. *Lancet* **379**, 713-720. doi:10.1016/S0140-6736(12)60028-2

Schwartz, S. D., Regillo, C. D., Lam, B. L., Eliott, D., Rosenfeld, P. J., Gregori, N. Z., Hubschman, J.-P., Davis, J. L., Heilwell, G., Spirn, M. et al. (2015). Human embryonic stem cell-derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow-up of two open-label phase 1/2 studies. *Lancet* **385**, 509-516. doi:10.1016/S0140-6736(14)61376-3

Sen, S. Q., Chanchani, S., Southall, T. D. and Doe, C. Q. (2019). Neuroblast-specific open chromatin allows the temporal transcription factor, Hunchback, to bind neuroblast-specific loci. *eLife* **8**, e44036. doi:10.7554/eLife.44036

Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., Choi, K., Bendall, S., Friedman, N. and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nat. Biotechnol.* **34**, 637-645. doi:10.1038/nbt.3569

Sieweke, M. H. (2015). Waddington's valleys and captain Cook's islands. *Cell Stem Cell* **16**, 7-8. doi:10.1016/j.stem.2014.12.009

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. doi:10.1016/j.stem.2011.07.014

Später, D., Hansson, E. M., Zangi, L. and Chien, K. R. (2014). How to make a cardiomyocyte. *Development* **141**, 4418-4431. doi:10.1242/dev.091538

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E. and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477. doi:10.1186/s12864-018-4772-0

Tanabe, K., Ang, C. E., Chanda, S., Olmos, V. H., Haag, D., Levinson, D. L., Südhof, T. C., and Wernig, M. (2018). Transdifferentiation of human adult peripheral blood T cells into neurons. *Proc. Natl. Acad. Sci. USA* **115**, 6470-6475. doi:10.1073/pnas.1720273115

Taïhi, I., Nassif, A., Isaac, J., Fournier, B. P. and Ferré, F. (2019). Head to knee: cranial neural crest-derived cells as promising candidates for human cartilage repair. *Stem Cells Int.* **2019**, 9310318. doi:10.1155/2019/9310318

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676. doi:10.1016/j.cell.2006.07.024

Takahashi, K. and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183-193. doi:10.1038/nrm.2016.8

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S. et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72-78. doi:10.1038/s41586-018-0654-5

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381-386. doi:10.1038/nbt.2859

Tsunemoto, R., Lee, S., Szűcs, A., Chubukov, P., Sokolova, I., Blanchard, J. W., Eade, K. T., Bruggemann, J., Wu, C., Torkamani, A. et al. (2018). Diverse reprogramming codes for neuronal identity. *Nature* **557**, 375-380. doi:10.1038/s41586-018-0103-5

Tursun, B., Patel, T., Kratsios, P. and Hobert, O. (2011). Direct conversion of C. elegans germ cells into specific neuron types. *Science* **331**, 304-308. doi:10.1126/science.1199082

Velasco, S., Ibrahim, M. M., Kakumanu, A., Garipler, G., Aydin, B., Al-Sayegh, M. A., Hirsekorn, A., Abdul-Rahman, F., Satija, R., Ohler, U. et al. (2017). A multi-step transcriptional and chromatin state cascade underlies motor neuron programming from embryonic stem cells. *Cell Stem Cell* **20**, 205-217.e8. doi:10.1016/j.stem.2016.11.006

Velasco, S., Kedaigle, A. J., Simmons, S. K., Nash, A., Rocha, M., Quadrato, G., Paulsen, B., Nguyen, L., Adiconis, X., Regev, A. et al. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523-527. doi:10.1038/s41586-019-1289-x

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. doi:10.1038/nature08797

Wang, N., Zheng, J., Chen, Z., Liu, Y., Dura, B., Kwak, M., Xavier-Ferrucio, J., Lu, Y.-C., Zhang, M., Roden, C. et al. (2019). Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nat. Commun.* **10**, 95. doi:10.1038/s41467-018-07981-6

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. doi: 10.1016/j.cell.2013.09.028

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. and Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, eaaw3381. doi:10.1126/science.aaw3381

Wolf, F. A., Angerer, P. and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15. doi:10.1186/s13059-017-1382-0

Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L. and Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59. doi:10.1186/s13059-019-1663-x

Xiao, D., Liu, X., Zhang, M., Zou, M., Deng, Q., Sun, D., Bian, X., Cai, Y., Guo, Y., Liu, S. et al. (2018). Direct reprogramming of fibroblasts into neural stem cells by single non-neural progenitor transcription factor Ptf1a. *Nat. Commun.* **9**, 2865. doi:10.1038/s41467-018-05209-1

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J. A. and Kosik, K. S. (2009). MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell* **137**, 647-658. doi:10.1016/j.cell.2009.02.038

Xu, Z., Jiang, H., Zhong, P., Yan, Z., Chen, S. and Feng, J. (2016). Direct conversion of human fibroblasts to induced serotonergic neurons. *Mol. Psychiatry* **21**, 62-70. doi:10.1038/mp.2015.101

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R. et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917-1920. doi:10.1126/science.1151526

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., Mcdermott, G. P., Zhu, J. et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049. doi:10.1038/ncomms14049

DEVELOPMENT