universität wien

DISSERTATION | DOCTORAL THESIS

Titel | Title

Functional and genetic dissection of transcription factor-mediated neuronal conversions from embryonic stem cells

verfasst von | submitted by

Gintautas Vainorius BSc MSc ETH

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 620 490
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von | Supervisor:	Ulrich Elling

Audrūnei

Acknowledgments

First and foremost, I would like to thank **Uli** for his guidance, support, and for fostering an environment where curiosity and rigor could thrive. To the entire lab, thank you for the collaborative spirit and turning the shared frustrations into the small victories along the way. To the **Negronis** — the legendary PhD selection, arguably the most sarcastic and humorous scientific cohort ever assembled — thank you for the laughter, the wit, and the shared madness. To my students, whom I had the privilege to teach — thank you for challenging me. I am proud of all that you have achieved.

Naomi - you were the best and closest friend I could have ever wished for. In the darkest moments, you helped me to stand up and take the next step; in the brightest, I reached twice as far because of you. **Maria and Georg**, thank you for finding me and dragging me along from the first day, for creating enthusiastic lab atmosphere filled with intellectual sparring, and for your lasting friendship afterwards. **Fanny**, thank you for joining me on adventures that reminded me that there's life beyond the lab. A special thanks to **Chris**, who was supportive and broke the protocols of the professionalism standards.

Lastly, I want to thank my family for your love, unwavering support and patience. **Justina and Audrūne** - thank you for waiting, for understanding, for being resilient in the face of my long absences. To **my brother and my parents** - you always were my biggest inspirations to whom I look up to.

Table of Contents

Summary

Over the past decades, ectopic expression of lineage-specifying transcription factors (TFs) has enabled direct cell fate conversion, offering new insights into cellular identity change and maintenance, and potential for regenerative medicine. Despite significant advances in the field, several fundamental issues remain unresolved. A wide array of TFs and their combinations have been shown to be able to instruct a new cell fate, yet it is unclear whether certain TFs are inherently more effective or mechanistically advantageous than others. The direct implications of TF choice on reprogramming efficiency, fidelity, and safety remain largely unexplored. Furthermore, it is also uncertain if these transitions occur via direct conversion, involve intermediate states or produce alternative lineages, some of which may be undesirable or even detrimental. Finally, the mechanisms by which different TFs orchestrate the transitions between cell states remain elusive, limiting the rational design of reprogramming strategies.

In this study, I focused on two proneural basic helix loop helix TFs, Ascl1 and Neurogenin2 (Ngn2), both capable of converting mouse embryonic stem cells (mESC) into induced neurons (iN). Besides sharing structural and developmental role similarities, they activate only partially overlapping transcriptional response in mESC. Expression of Ascl1 rapidly dismantles the pluripotency network and installs a new cellular fate, but also induces a trophoblast-like cell population, likely due to its similarity to Ascl2, a trophoblast specific TF. In contrast, Ngn2 initiates a more interconnected transcriptional response, with cells transitioning through a neural stem cell-like intermediate, supported by an incomplete shutdown of the pluripotency network. CRISPR-Cas9 knockout screen revealed that Ascl1, but not Ngn2, relies more on factors regulating pluripotency and cell cycle, such as a transcriptional repressor Tcf7l1. Upon Tcf7l1 knockout, cells fail to upregulate Cdkn1c, exit cell cycle, and establish iN state. However, overexpression of Cdkn1c induces cell cycle arrest and restores iN conversion.

These results demonstrate that cell type conversion can proceed through distinct mechanistic paths, despite being driven by closely related transcription factors. Each path is characterized by unique transcriptional and functional properties and distinct side populations, underscoring the importance of understanding the processes underlying cell fate transitions.

Zusammenfassung

In den letzten Jahrzehnten hat die ektopische Expression von lineagespezifischen Transkriptionsfaktoren (TFs) die direkte Umprogrammierung von Zellschicksalen ermöglicht. Dies hat neue Einblicke in die Plastizität zellulärer Identität sowie vielversprechende Perspektiven für die regenerative Medizin eröffnet. Trotz bedeutender Fortschritte bleiben jedoch grundlegende Fragen unbeantwortet. Eine Vielzahl von TFs und deren Kombinationen kann neue Zellschicksale induzieren, doch ist unklar, ob bestimmte TFs inhärent effektiver oder mechanistisch vorteilhafter sind als andere. Die direkten Auswirkungen der TF-Wahl auf Reprogrammierungseffizienz, -treue und -sicherheit sind bislang weitgehend unerforscht. Zudem ist ungewiss, ob diese Übergänge durch direkte Umwandlung erfolgen, Zwischenzustände durchlaufen oder alternative Zelllinien erzeugen – von denen einige unerwünscht oder sogar schädlich sein könnten. Auch die Mechanismen, durch die verschiedene TFs den Übergang zwischen Zellzuständen steuern, sind weitgehend unbekannt, was die rationale Entwicklung von Reprogrammierungsstrategien erschwert.

In dieser Studie habe ich mich auf zwei proneurale basic-helix-loop-helix-TFs konzentriert: Ascl1 und Neurogenin2 (Ngn2), die beide in der Lage sind, murine embryonale Stammzellen (mESCs) in induzierte Neuronen (iNs) umzuwandeln. Obwohl sie strukturelle Ähnlichkeiten und vergleichbare entwicklungsbiologische Funktionen aufweisen, aktivieren sie in mESCs nur teilweise überlappende transkriptionelle Programme. Die Expression von Ascl1 führt rasch zur Auflösung des Pluripotenznetzwerks und zur Etablierung eines neuronalen Schicksals, erzeugt jedoch zusätzlich eine trophoblastenähnliche Zellpopulation – vermutlich aufgrund seiner Ähnlichkeit zu Ascl2, einem trophoblastenspezifischen TF. Im Gegensatz dazu initiiert Ngn2 eine stärker vernetzte transkriptionelle Antwort, wobei die Zellen über einen neuralen Stammzell-ähnlichen Zwischenzustand differenzieren, unterstützt durch eine unvollständige Abschaltung des Pluripotenznetzwerks. Ein CRISPR-Cas9-Knockout-Screen zeigte, dass Ascl1 – im Gegensatz zu Ngn2 – stärker auf Faktoren angewiesen ist, die Pluripotenz und Zellzyklus regulieren, wie etwa den Transkriptionsrepressor Tcf7l1. Nach Knockout von Tcf7l1 gelingt es den Zellen nicht, Cdkn1c hochzuregulieren, den Zellzyklus zu verlassen und den iN-Zustand zu etablieren. Die Überexpression von Cdkn1c hingegen führt zum Zellzyklusarrest und stellt die neuronale Umwandlung wieder her.

Diese Ergebnisse verdeutlichen, dass Zellschicksalsumwandlungen selbst bei eng verwandten Transkriptionsfaktoren über unterschiedliche mechanistische Pfade verlaufen können. Diese Pfade sind durch einzigartige transkriptionelle und funktionelle Eigenschaften sowie durch unterschiedliche Nebenpopulationen gekennzeichnet, was die Bedeutung eines tieferen Verständnisses der zugrunde liegenden Prozesse bei Zellschicksalsübergängen unterstreicht.

Introduction

Understanding lineage transitions – a historic perspective.

In 20th century Conrad Hal Waddington devised a model depicting embryonic development as a ball rolling downhill along a branching landscape. As the ball progresses, it makes incremental choices that progressively restricts its potential, ultimately guiding it towards the base of the valley – a terminally differentiated cell state. Indeed, decades of careful *in vivo* and cell culture differentiation experiments showed that cells gradually rewire their gene transcriptional network, reshape their epigenetic landscape, adapt metabolism and cellular morphology, and establishing epigenetic barriers to lock the cellular identity [1] . However, the directionality of cell fate transitions was challenged by Sir John Gurdon experiments of somatic nuclear transfer, where nuclei from terminally differentiated cells from Xenopus tadpole's intestinal epithelium cells were placed into enucleated oocytes [2] . After transplantation, oocytes could develop and give rise to viable tadpoles with identical genetic materials as donor tadpoles. Years later, similar experiments were repeated to clone mammalian organisms, including mice, dogs, cows, and sheep [3,4] , as well as reverting human fibroblast to embryonic stem cell state [5] . Together, these experiments proved that somatic cells possess a capacity to revert to early undifferentiated state, seemingly inverting the Waddington model, and reactivate the full program of the embryonic development.

An interplay of the transcription factors (TF) acts as a primary driver for cell lineage transitions and establishment of cellular identity [6,7] . As such, overexpression of lineage specifying TFs allowed conversions from one cell type to another. In the earliest example, Myod1, a TF governing differentiation and maintenance of skeletal muscle cells, was able to impose myoblasts cell fate when ectopically expressed in mouse fibroblasts [8] . In a breakthrough study, Shinya Yamanaka have shown that Oct4, Sox2, Klf4, c-Myc (OKSM) [9,10] , a cocktail of TF that are present in mouse embryonic stem cells (ESC), can readily reprogram mouse embryonic fibroblast (MEF), back to pluripotency state as induced pluripotent stem cells (iPSC), which subsequently can be differentiated into practically any cell type. Notably, same factors are also able to reprogram multiple cell types [11] , as well as generate iPSC from multiple species, including human cells [10] . Most importantly, cells derived via cellular reprogramming are transcriptionally and functionally indistinguishable from their developmental counterparts, as iPSC cells can integrate into preimplantation embryos, subsequently contributing to normal development, and even produce viable mice in tetraploid complementation assays [12] .

Cell identity conversions are not constrained within the same germ layers or reversal to pluripotency state. New cell identity can be instructed between distant cell types across germ layers, in a process termed transdifferentiation. In a landmark study, MEF could be converted to induced neurons (iN) via

overexpression of pro-neural TF Ascl1, in combination with Brn2 and Myt1l (BAM) [13,14] . Additionally, same factors are able to reprogram other cell types to iN e.g. hepatocytes, skeletal muscle myoblasts, osteoblasts to name but a few [15,16] . Similarly, Ascl1 can convert ESC to terminally differentiated neurons in a rapid and more efficient manner, termed directed differentiation [17] . Importantly, conversion between two cell identities via overexpression of lineage specifying TFs ignores developmental trajectories as well as developmental intermediate cell states in both transdifferentiation and directed differentiation paradigms [18,19] .

More recently cellular reprogramming methods proved to be a major advance in therapeutic and disease modeling applications. For example, somatic cell conversion to iPSC allowed to study patient specific disease phenotypes and can be used in cellular replacement therapies [20]. Additionally, direct conversion paradigm allows for *in situ* conversion of cells within the tissue, offering a promising strategy of repairing damaged tissues or replenishing functional cell pool without the need for cell transplantation, such as in neurogenerative diseases or myocardial infarction [21,22]. Lastly, similarities have been observed between lineage conversion mechanism and pathological transformations e.g. tumorigenesis[23] , providing multiple means in understanding and probing constrains of cell plasticity. Thus, a detail understanding of these mechanisms and improving induced lineage conversion methods are essential for in depth understanding of pathologies as well as developing new and safe therapeutic modalities.

In this section I will give a brief overview about mechanistic understanding and advances in cellular reprogramming, with a specific focus on cell lineage conversions towards iN and iPSC.

Ascl1 and Ngn2 in development and reprogramming.

Ascl1 (also known as mammalian achaete-scute homolog 1, Mash1) and Ngn2 (encoded by gene *Neurog2*) are basic helix-loop-helix (bHLH) family transcription factors and are master regulators of neurogenesis, essential in development of the central and peripheral nervous systems.

Ascl1 and Ngn2 is expressed early in neural development during the formation of the neural tube and is crucial for the differentiation of neural stem cells (NSC) into neurons and repressing differentiation towards gliogenesis [24] . Oscillatory expression of Hes proteins, effectors of the Notch pathway, are required for maintenance of NSC state [25,26] . However, Hes proteins repress proneural bHLH gene expression, including Ascl1 and Ngn2, hence driving opposite phased oscillations of Ascl1 and Ngn2 [25,27] . In turn, Ascl1 and Ngn2 activate the expression of the Notch ligand Delta-like 1 (Dll1), which leads to the downregulation of Hes proteins via lateral inhibition [25] . Sustained expression of Ascl1, Ngn2 or repression of Hes proteins leads to the exit of NSC state and neuronal differentiation [28] .

Interestingly, Ascl1 has two opposing roles in the neurogenesis: promoting proliferation of progenitors and initiating their differentiation and cell cycle exit [29] . Ascl1 directly binds and regulates both cell cycle exit genes and genes important for G1/S transition and entry into mitosis [29] . Acute loss of Ascl1 results in premature cell cycle exit in NSC [29] . Notably, different Ascl1 expression dynamics was shown to govern cell cycle regulation: oscillatory Ascl1 expression promotes proliferation of NSC, while sustained expression instead drives differentiation [26] .

Within telencephalon, Ascl1 is expressed primarily in ventral progenitor cells, instructing them to adapt GABAergic neuronal fate [30,31] . In contrast, Ngn2 is expressed in ventricular zone (VZ) and subventricular zone (SVZ) of dorsal cortical progenitors, directing them toward glutamatergic neuronal fate [30,31] . Despite their distinct roles in specifying different neuronal subtypes in the telencephalon, expression of Ngn2 in the ventral telencephalon can rescue Ascl1-null mice [30] .Alternatively, Ascl1 expressed in dorsal telencephalon respecifies neuronal progenitors towards GABAergic neuronal fate [30,32] . This suggests that Ngn2 has as a permissive role, acting in combination with other factors, while Ascl1 functions as an instructive factor in specifying neuronal subtypes.

Lastly, Ascl1 aslo plays a pivotal role in adult neurogenesis in rodents [33,34] . In hippocampus, expression of Ascl1 in quiescence adult NSCs promotes their activation and differentiation to neurons, while conditional knockout of Ascl1 leads to inability of adult NSCs to exit quiescence and stay unresponsive to the activation stimuli [33,34] .

Roadblocks and hurdles in cellular reprogramming

During differentiation towards a terminal identity, cells progressively restrict genes expression through multiple layers of epigenetic regulations. These include deposition of repressive histone marks, such as H3K27me3 by Polycomb complex, the formation of heterochromatin enriched in H3K9me3, promoter DNA methylation by DNMTs at CpG-rich regions, as well as genomic elements can be spatially distanced and restricted during three-dimensional chromatin reorganization [1,35–37] . In turn, cells upregulate coherent autoregulatory gene network that reinforce and maintain cellular identity [1] . This ensures a stable differentiated cells state and prevent aberrant gene activation interfering with cellular function as well as preventing cells from pathological and cancerous transformations [1,38–40] . Indeed, many cancers are shown to have dysregulation of histone marks deposition and DNA methylation [1,40] . Furthermore, aberrant chromatin landscape is prone to activation of unwanted genetic programs, for example partial reactivation of pluripotency or progenitor transcriptional programs in numerous cancers [41] . It is therefore striking that ectopic expression of transcription factors outside of their developmentally defined context can effectively navigate these epigenetic roadblocks and unlock

repressed genome regions in a controlled and reproducible manner, resulting in novel non-developmental conversion trajectories.

Nucleosomes provide a physical barrier for the transcription factor accessibility for induction target program. However, Ascl1 induces a coordinated and rapid genome-wide chromatin remodeling already 12h post induction and 80% of all accessibility changes occur rapidly within 2 to 5 days [42] . This chromatin switch results in nucleosome phasing around neuronal gene promoters and enhancers resulting in stable genetic neuronal program expression [42] . Similarly, Ascl1 and Ngn2 induced increased chromatin accessibility at binding sites upon expression in ESC after 12h post induction [43,44] . Newly accessible regions can then be occupied by downstream transcription factors [43,44] . Interestingly, between Ascl1 and Ngn2, only Ascl1 alone is able to convert MEF to iN [13] , while Ngn2 is able to bind and partially induce neuronal program, but fails to generate iN [45,46] . However, addition of small molecules Forskollin and Dorsomorphin increases chromatin accessibility as well as H3K27 acetylation, leading to induction of SOX4, NEUROD1 and NEUROD4, and highly efficient iN conversion [46] .

Histone H3K9 di- and tri-methylation is found in constitutively repressed heterochromatic regions and is increasingly deposited during cell fate changes to restrict the lineage specific gene expression [47,48] . Hence, cellular reprograming needs to reactivate target lineage genes “locked” in heterochromatin. For example, core pluripotency genes in somatic cells are in H3K9me3-enriched chromatin regions and required to be reactivated during iPSC reprogramming [49] . This restricts access for pluripotency factors and their activation is essential for successful iPSC reprogramming [49,50] . Not surprisingly, knockdown of histone H3K9 methyltransferases SUV39H1/H2, implicated in heterochromatin formation, reduced H3K9me3 levels and allowed a more robust and efficient reprogramming [49,51] . Similarly, heterochromatic H3K9me3 marked regions impeded activation of liver specific genes during human fibroblast to induced hepatic cells transdifferentiation [52] . Suppression of SUV39H1 or RBMX and RBMXL1, H3K9me3 readers, relieved hepatic gene suppression improving efficiency of induced hepatic cell conversion [52] . Thus, H3K9 methylation acts as general safeguard restricting expression of unwanted genetic programs and stabilizing cell identity. Surprisingly, Ascl1 in MEF favors regions marked by trivalent histone marks, consisting of H3K9me3 H3K4me1 and H3K27ac, for efficient binding and activation of marked target genes[16] . This trivalent chromatin state can predict Ascl1 binding as well as efficiency of iN conversion between different cell types [16] . Expression of JmjD2d H3K9me3 histone demethylase prior to Ascl1 induction led to reduction of global H3K9me3 levels and the reduction of generation of iN [16] .

Outside unfavorable chromatin landscape, transcriptional repressors prevent activation of unwanted alternative lineages[53] . The REST is a transcriptional repressor of neuronal genes in non-neuronal cells

by binding to the RE1 sites, also known as neuron-restrictive silencer elements (NRSE) [54] . Thus, during iN conversion, the REST complex needs to be inhibited to allow for the activation of neuronal genes that are typically suppressed in non-neuronal cells. Indeed, astrocytes becomes refractory to iN conversion by Ngn2 or, to a lesser extent, Ascl1 due to REST activity preventing binding and the induction of downstream targets, such as NeuroD4, and inhibition of REST complex enhances iN conversion [55] .

Lastly, cell cycle can play a role in both facilitating and inhibiting lineage conversions. Proliferation shown to play an important role in MEF reprogramming to iPSC, likely allowing fast erasure of epigenetic marks, such as DNA methylation [56] . Although c-Myc is dispensable from OKSM TF cocktail, c-Myc expression increases cell cycle leading to a more efficient iPSC generation [57,58] . In contrast, neurons are postmitotic and failure to stop cell cycle can lead to neurodevelopmental disorders, such as microcephaly or megalencephaly[59,60] . Thus, in contrast to iPSC reprogramming, MEF to iN conversion can happen without cell division and cell cycle is rapidly stopped after Ascl1 induction [61,62] , while inhibiting proliferation improves fibroblast to neuron reprogramming[62,63] .

Finally, during DNA replication in cycling cells, chromatin is disassembled, and region-specific histone marks need to be restored to maintain epigenetic memory [64] . The epigenetic structure of chromatin is preserved through nucleosome repositioning by ATP-depended chromatin assembly complexes, such as CAF-1, thus reestablishing epigenetic roadblocks for ectopically expressed TFs. In a genetic siRNA screen, it was found that depletion of CAF-1 subunits Chaf1a or Chaf1b resulted in a more accessible chromatin, reduction of heterochromatin domains and increased binding of Sox2 to the pluripotency genes during MEF to iPSC reprogramming[65] . Interestingly, inhibition of CAF-1 also enhanced conversion of MEF to iN by Ascl1 overexpression and B cells to macrophages following expression of C/EBPα, showing histone turnover can act as a broad mechanism for cell fate control [65] . Likewise, during DNA replication binding of cell lineage TFs must be reestablished [66] . This creates a window of opportunity for competing lineage drivers to have an engage a loosen chromatin landscape. Thus, not surprisingly, Oct4, Klf4 and Sox2 were shown to be able to bind during cell division and possess “bookmarking” TF properties [66] .

Pioneering factors as a choice for lineage conversions

Only a small fraction of genome is open and actively transcribed, while majority consists of quiescent, low signal chromatin or closed, transcriptionally silent chromatin marked with H3K27me3 and H3K9me3 repressive histone marks [67] . These silent heterochromatic regions present a barrier for the TF accessibility due to the high DNA-binding affinity of nucleosomes, which sequester genes essential for the target cell state [68,69]. However, a subset of TFs, termed “pioneer factors”, are able to bind and

open closed chromatin by partially recognizing DNA motifs wrapped around nucleosomes [69] . Binding of target sites even in the presence of a nucleosome, subsequently rearranging accessibility of chromatin and inducing gene expression allows pioneering factors to drive cell fate changes in development [16,68,70] . Given their ability to access and remodel closed chromatin, pioneer factors are frequently selected for cellular reprogramming strategies as drivers or to enhance efficiency and specificity.

Early studies identified FoxA and GATA4 as capable of binding liver-specific *alb1* enhancer well before activation of gene expression during liver development in early mouse embryos [71] . Consequently, FoxA and GATA TFs are widely utilized in reprogramming towards hepatocytes [15,72–74] . In iPSC reprogramming, Oct4, Klf4, Sox2 are engaging multiple genes, including pluripotency network genes, marked by H3K9me3 [49] . However, their binding is highly influenced by co-expression and cooperative binding between the OKSM factors and changes dynamically during different reprogramming stages [75,76] . Hence, OKS binding is also context dependent and binding in chromatin landscape dependent manner. Of note, c-Myc does not possess pioneering activity and targets preferentially accessible sites at promoters marked by activating histone modifications, and later new accessible sites opened by OKS facilitates c-Myc binding [49] . In contrast, Ascl1 exhibits “on-target pioneering” binding activity largely independent of chromatin context and binds same targets in fibroblasts and neuroblastoma cells as in NSC[16,77] . Similarly, after 48h of Ascl1 induction many Ascl1 binding sites are also shared between mouse ESC and MEF, even though initial binding of Ascl1 differs between these cell types [43,78] . Upon binding, Ascl1 then induces opening of chromatin and allowing recruitment of downstream effectors, such as Brn2, facilitating transdifferentiation to iN [16,70] .

Molecular mechanism of cell fate reprogramming

A key difference from the classical developmental view of cellular identity is that cell lineage reprogramming disregards directionality and trajectories of the developmental genetic program. For example, iPSC reprogramming erases somatic cell identity and reinstalls early pluripotency state [79]. Transdifferentiation allows cells to cross borders of germ layers without pluripotent intermediates e.g. MEF to iN conversion [14,15] . Directed differentiation, on the other hand, bypasses multiple developmental stages while able to establish alternative intermediate states [18,19] . Thus, ectopic expression allows cells to transition via different points in Waddington landscape, suggesting that cell states can be viewed independently of their derivation history, that is, regardless of the originating cell type or direction a new cellular state was established. In such, cellular states act as multidimensional attractors, exhibiting local minima of all possible stable genetic network states, akin “valleys” in Waddington landscape. Such attractors can be defined by a set of molecular parameters, such as gene expression, protein concentrations, sensitivity to signaling molecules, and more[80,81] . Lastly, cells can

transition between or return to initial state, after stimulus or small perturbations, respectively [80–82] . Most importantly, these transitions are not random but follow highly reproducible and defined molecular logic.

Ascl1 induced MEF to iN direct conversion happens in a hierarchical stepwise manner [16,42]. Due to pioneering activity, Ascl1 directly binds and activates neuronal specific genes, inducing rapid chromatin rearrangement within the first few days of transdifferentiation, followed by a fast downregulation of fibroblast program and induction of iN program [16,42]. In MEF and mouse ESC, Ascl1 binding leads to an increase of activation histone mark H3K27ac and DNA accessibility at target sites, which in turn allows binding of downstream effectors, such as Brn2, to further solidify the induction of neuron-specific genes and solidify the induction of neuronal program[16,42,43] . Similarly, Ngn2 together with Isl1 and Lhx3 converts in ESC to induced motor neurons in a stepwise transcriptional feedforward logic manner [70]. Initially, Ngn2 and Isl1/Lhx3 binds distinct regulatory regions. However, Ngn2 induced TFs Ebf2 and Onecut2 TF subsequently direct binding of Isl1/Lhx3 to the motor neuron-specific enhancers during later stages of differentiation, thereby stabilizing terminal neuronal fate [70] .

Although Ascl1 is an "on-target" pioneering factor, however, expressed in MEF, Ascl1 binding partially overlaps with binding sites of Myod1, a muscle lineage inducing bHLH transcription factor [77,78] . Subsequently, a myogenic transcriptional program is induced alongside neuronal program, competing for the outcome of the conversion [77,78,83] . To counteract this, Myt1l is commonly included in the TF cocktails for the iN conversion[14] . Surprisingly, Myt1l does not bind and induce neuronal genes, however, it acts as an "all-but-neuronal-state" repressor and is important in repressing muscle lineage, fibroblast and other donor cell programs [84] . Additionally, it was shown to be required for the efficient exit of the neural progenitor state during development[84] . Interestingly, similarly to Ascl1, Myod1 also binds neuronal targets when expressed in MEF [78] . Consequently, co-expression of Myod1 with Myt1l leads to the formation of iN [78] , showing the importance of studying in detail the mechanism of action of the conversions and carefully crafting the transcription factor cocktails for therapeutic applications.

In contrast, MEF to iPSC reprogramming consists of 2 phases – stochastic and hierarchical[50] . Initially, OKS directly bind and decommission fibroblast program specific enhancers while gradually activating pluripotency related enhancers [75] . However, OKSM factors engage genome and activate target gene expression with a high degree of variability between cells during early stages of reprogramming [50] . Furthermore, many pluripotency network (PPN) genes, such as *Sox2*, *Nanog* and *Prdm14*, are tightly packed and repressed as part of heterochromatin and are refractory from initial activation of OKSM factors[49]. Activation of these genes, especially *Sox2*, corresponds to a hierarchical stage of

reprogramming characterized by a predictive sequence of gene activation leading towards pluripotency [50] .

Cellular reprogramming in understanding principles of lineage specification

Ability of transitioning between cell states outside development provides an exciting experimental framework to address principles of cell fate specification, maintenance and safeguarding of cell identity, where cells are able to both exit and enter different terminal states via multiple routes. For this several strategies can be employed.

(1) Cells can exit initial state in multiple ways:

Thus, multiple distinct mechanisms can be studied to understand how somatic cell state can be compromised. For instance, MEFs can be reprogrammed to iPSC by OKSM or to iN by Ascl1. iPSC reprogramming is dependent and greatly improved by increased cell proliferation and inhibited by cell cycle arrest and cellular senescence [85,86] . In contrast, MEF to iN conversion by Ascl1 can happen without cell cycle and inability to exit cell cycle leads to decrease in transdifferentiation efficiency [61,62] . Alternatively, Francesconi M. and colleagues reprogrammed mouse pre-B cells into iPSC cells using OKSM or transdifferentiated into macrophages by expression of C/EBPα and showed that high c-Myc expression in the initial population correlates with the better iPSC reprogramming efficiency but worse macrophages conversion [87] .

(2) Cells can be converted to desired cell type from different host cell types:

OKSM can universally reprogram multiple somatic cell types back to iPSC state [11] , while BAM factors were shown to be sufficient to convert various cell types to iN [16] . This suggest that both OKSM and BAM TF cocktails possess properties allowing them to dismantle broad range of host transcriptional networks and install target cell transcription network regardless of initial cell identity. Indeed, unique features in genome binding, activity on gene expression, and cooperativity have been described for the OKSM and BAM (see above) [13,16,49,66,69,75,88] . Furthermore, this approach can provide mechanistic requirements for establishing target cell identity as well as elucidating cell type specific effects [11] .For example, iPSC reprogramming trajectories from fibroblasts, neutrophils, and keratinocytes converge onto iPSC state via two transcriptional waves, original cell identity loss and pluripotency network establishment. However, they are significantly influenced by host cell transcriptional networks dictating the speed, stages and intermediates during reprogramming [11] .

(3) Cells can transition between two cell states via different paths:

Multiple TF cocktails exist allowing to convert cells between two states, which must induce factors required for target cell type conversion and maintenance. Furthermore, a set of genetic roadblocks

exists between two states (see above) which TFs cocktails need to navigate through via similar or different pathways. For example, the astrocytes can be converted to iN by overexpression of Ascl1 or Ngn2 [55] . Even though both TFs produce rapid changes in gene expression upon overexpression, and rapidly generate iN with high efficiency, only a fraction of upregulated genes are common between them. Several downstream factors, namely Neurod4, was sufficient in replacing Ascl1 or Ngn2 for iN conversion [55] . Briggs et al., 2017, compared two motor neuron differentiation protocols: standard *in vitro* differentiation protocol mimicking development and directed differentiation using ectopic expression of Ngn2, Isl1, Lhx3. Interestingly, while terminal states were similar, directed differentiation protocol bypasses multiple progenitor states defined in embryo, and produced a novel transition state with unique gene expression, which later converges to the final motor neuron state [18] . Similarly, Stuart et al., 2019 studied the reversal of epiblast stem cells (EpiSC) state back to Naïve ESC state upon induction of Esrrb, Klf2, two TFs in pluripotency network, or a Jak/Stat3 activating receptor GY118F [89] . Each transgene modulates different signaling axes effecting naïve pluripotency network: Wnt, Fgf, and LIF, respectively. Expression of all transgenes resulted in cells reprogramming via mechanistically and transcriptionally different trajectories albeit ending at the same terminal state. Surprisingly Klf2 expression leads to temporary induction of mesodermal genes before acquisition of naïve pluripotency. In these transitions, a unifying feature is a precise regulation of Oct4 expression, which is both required and sufficient for EpiSC to ESC reprogramming [89] .

(4) Same conversions but in cells from different organisms:

Same factors can drive the conversions of cells in different species, albeit different efficiencies. Ascl1 alone can convert mouse tail tip fibroblasts to mature neurons, however, only additional expression of Brn2 and Myt1l allows transdifferentiation of human dermal fibroblasts, suggesting a more robust genomic regulation in human cells [13,14] . In contrast, OKSM factors universally reprogram both mouse [9,90] , human cells [10,91] and many other mammalian cells [92] . A unique model organism in which reprogramming to iPSC has been investigated is the naked mole rats, a rodent known for its exceptional longevity and resistance to tumorigenesis. Interestingly, naked mole rat's embryonic fibroblasts require overexpression of SV40 Large T antigen or inactivation of retinoblastoma protein Rb to overcome a unique epigenetic stability of the cells [93,94] . Similarly to the species remarkable cancer resistance, iPSC derived from naked mole rats have a very low propensity of teratoma formation [93,95] . This may be attributed to the high levels of H3K27/K36 methylation, corresponding to a repressive heterochromatin, as well as a more closed chromatin around reprogramming important genes. These features result in a more stable epigenome that resists reprogramming compared to mouse and human cells [93] . Thus, studying different interspecies cell conversion mechanisms can illuminate different layers of epigenetic stability and cell identity maintenance.

Given many possibilities to transition via multiple cell identities, a tempting question to ask is if common rules and mechanisms exist between multiple lineage transitions or whether every conversion is a unique process linking two cell identities. Likewise, what properties between different sets of TF are important for both upregulation of target gene regulatory network (GRN) and exit of the host cell state, and whether different strategies can be employed. Lastly, identifications of roadblocks will lead to better understanding of cellular identity stabilization and maintenance as well as improving efficiency of reprogramming methods for therapeutic applications

Advances in disease modeling and therapeutic applications.

Lineage conversion paradigm has emerged as a novel tool in the applied research and regenerative medicine, offering new angles for exploring underlying disease mechanisms and designing effective treatment strategies for a wide range of human pathologies.

The discovery of OKSM factors allowed conversion of patient-specific somatic cells to iPSC, which can subsequently be differentiated into disease-relevant cell types. This approach offers significant advantages over established cell lines or animal models, as it allows modeling of the disease onset, progression and patient specific aspects of diseases mechanism. Notably, iPSC reprogramming erases epigenetic and aging related hallmarks, resulting in decreased DNA methylation and expression of age-associated genes, improved nuclear envelope integrity and mitochondrial function, and increased telomere length [96,97] . However, direct conversion of fibroblasts to neurons retains age specific traits [98–100] and can be used to study aging related diseases such as Parkinson's, Alzheimer, Huntington diseases as well as Amyotrophic lateral sclerosis. Furthermore, lineage conversion can be used to derive cell lines from large cohort of people to account for complex traits effecting the mechanism, like gender, age, ethnicity, and disease severity. These cell line libraries can be then used for drug screening, assessing efficacy, potential adverse or side effects in *in vitro* assays to better reflecting patient population or in a patient specific manner.

Multiple pathologies result in the loss of cells, damage to the tissue and failure of the organ function. As such, direct lineage conversions are of particular use in clinical applications for patient-specific cell replacement therapies. These therapies can additionally be combined with gene editing to correct underlying genetic causes of the pathology prior to introducing cells back to the patients. For instance, the degeneration of dopaminergic neurons is a hallmark of Parkinson's disease [101] . Replenishing dopaminergic neuron pools has seen encouraging outcomes, leading to the long-term alleviation of symptoms in animal models and, more importantly, in patients [101] . Similar approaches are being explored for Alzheimer's disease [102], Huntington's disease and retinal degeneration [22] . Likewise, type 1 diabetes is a chronic autoimmune disease in which the body's immune system attacks and destroys

the insulin-producing β cells in the pancreas, leading to insulin insufficiency. A direct lineage conversion of non-β-cells, for example α-cells in pancreas, to insulin secreting glucose-responsive β-like cells, restores insulin production and normalizes blood glucose levels in animal models [103–105] .

A particularly attractive use of direct reprogramming in clinical applications is the *in situ* conversion of the endogenous cells to replenish cell pools in the diseases or after an injury. Astrocytes, which are abundant in the brain and reactive after injury, have been a primary target for direct conversion into functional neurons *in situ* [102] . In spinal cord injuries, permanent neuronal loss and the formation of glial scars leads to the long-lasting neurological impairments. *In situ* replacement of astrocytes or implantation of directly converted fibroblasts was shown to restore damaged neural circuits and improve of locomotor functions in animal models [106,107]. Similarly, cardiac injuries, such as myocardial infarction can result in the over-proliferation of cardiac fibroblasts, leading to the replacement of cardiomyocytes, collagen rich scarring, and impediment of the normal function of the heart [108] . Indeed, *in vivo* conversion of cardiac fibroblast to cardiomyocyte-like cells by Gata4, Mef2c, Tbx5 reduced tissue fibrosis and improved cardiac function in animal models [109,110] .

Although lineage conversion methods are increasingly being used in the applied research, several limitations exist and need to be addressed in future experiments. A purity of terminal populations should be always addressed in detail. For example, while iPSC can theoretically generate large numbers of target cells, it is vital to ensure a complete differentiation, as residual iPSC can lead to the teratoma formations [111] . Similarly, in depth molecular analysis of reprogrammed iN transplantation or iN *in situ* reprogramming is needed to evaluate both cell autonomous and non-cell autonomous effects to mitigate risk of incomplete neuronal phenotype, generation of aberrant or toxic cell identities and ensuring long term functional integration and stability of reprogrammed cells [112] . Different TF combinations also need to be explored for highly efficient iN subtype specification [113–116] , as well as the choice of donor cell type, as reprogrammed cells can retain regional identity, which can influence their integration and function [115] . Thus, it is vital to develop new and better methods for both efficient conversion and outcome quantification for preclinical and therapeutic applications.

Aims of the Thesis

The ability of different TFs to induce same fate provides an opportunity to contrast mechanisms of cell identity conversion, including how TFs potency, specificity and interaction with the host cells gene network shapes the trajectory of the conversion. Identifying key mechanistic factors and roadblocks, as well as cellular intermediates and alternative states could inform a further TF choice for therapeutic applications. The goals of this thesis are to compare molecular mechanisms of Ascl1 and Ngn2 driven ESC to iN directed differentiation.

Specifically, the goals are:

1) Profile Ascl1 and Ngn2 binding and activation of transcripts.
2) Compare terminal, intermediate and alternative cell populations.
3) Develop CRISPR-Cas9 forward genetics screening tools for investigating genetic dependencies underlying cell fate transitions.
4) Determine differential genetic dependencies between Ascl1 and Ngn2 generation of iN.

This thesis contains two papers. Paper #1 (Vainorius et al. 2023) describes the mechanistically distinct conversion paths of Ascl1 and Ngn2 of ESC to iN, highlighting that closely related TFs can induce mechanistically distinct conversion paths. Paper #2 (Michlits et al. 2017) establishes a CRISPR-UMI technology, its application for investigating roadblocks of iPSC reprogramming and later applied in elucidating genetic dependencies of ESC to iN conversion in Paper #1.

nature communications

Article

https://doi.org/10.1038/s41467-023-40803-y

Ascl1 and Ngn2 convert mouse embryonic stem cells to neurons via functionally distinct paths

Received: 11 July 2022

Accepted: 9 August 2023

Published online: 02 September 2023

Check for updates

Gintautas Vainorius[1,2] ✉, Maria Novatchkova[1,3], Georg Michlits[1,2,5], Juliane Christina Baar[1], Cecilia Raupach[1], Joonsun Lee[1,2], Ramesh Yelagandula[1,6], Marius Wernig[4] & Ulrich Elling[1] ✉

Ascl1 and Ngn2, closely related proneural transcription factors, are able to convert mouse embryonic stem cells into induced neurons. Despite their similarities, these factors elicit only partially overlapping transcriptional programs, and it remains unknown whether cells are converted via distinct mechanisms. Here we show that Ascl1 and Ngn2 induce mutually exclusive side populations by binding and activating distinct lineage drivers. Furthermore, Ascl1 rapidly dismantles the pluripotency network and installs neuronal and trophoblast cell fates, while Ngn2 generates a neural stem cell-like intermediate supported by incomplete shutdown of the pluripotency network. Using CRISPR-Cas9 knockout screening, we find that Ascl1 relies more on factors regulating pluripotency and the cell cycle, such as Tcf7l1. In the absence of Tcf7l1, Ascl1 still represses core pluripotency genes but fails to exit the cell cycle. However, overexpression of Cdkn1c induces cell cycle exit and restores the generation of neurons. These findings highlight that cell type conversion can occur through two distinct mechanistic paths, even when induced by closely related transcription factors.

Cell identity in development is typically crafted by an interplay of multiple transcriptional activators and inhibitors successively restricting the lineage. Cells gradually rewire their gene transcriptional network, retarget multiple transcriptional factors to new loci, reshape their epigenetic landscape, and establish epigenetic barriers to lock the next developmental state[1–5]. In contrast, directed lineage conversions rely on the expression of transcription factors (TF) in alien cellular contexts that can be affected by different distributions of epigenetic marks[6,7], posttranslational modifications[8,9], interaction partners[10–13], and other factors[14,15]. This can affect the binding and activity of the TF. Conflicting or incomplete cues of lineage formation allow the formation of alternative lineages during directed lineage conversion regimes[16–19]. These conversions do not necessarily follow developmental trajectories, and can also skip intermediate cell states[20,21]. As such, they provide an opportunity to study alternative inroads to cell states. Thus, cellular reprogramming can be used as a powerful tool to address fundamental principles underlying lineage specifications and generate models for a study of various diseases as well as potential therapeutic modalities[22,23].

Basic helix loop helix (bHLH) TFs have proven to be efficient factors for lineage conversion. Various proneural factors, such as Ascl1, Ngn1, Ngn2, Neurod1 and Neurod4 have been utilized to convert

[1]Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Dr. Bohr-Gasse 3, Vienna BioCenter (VBC), 1030 Vienna, Austria. [2]Vienna Biocenter PhD Program, a Doctoral School of the University of Vienna and Medical University of Vienna, A-1030 Vienna, Austria. [3]Research Institute of Molecular Pathology (IMP), Campus-Vienna-BioCenter 1, Vienna BioCenter (VBC), 1030 Vienna, Austria. [4]Institute for Stem Cell Biology and Regenerative Medicine, Department of Pathology, Stanford University, Stanford, CA, USA. [5]Present address: JLP Health GmbH, Himmelhofgasse 62, 1130 Vienna, Austria. [6]Present address: Laboratory of Epigenetics, Cell Fate & Disease, Centre for DNA Fingerprinting and Diagnostics (CDFD), Uppal, Hyderabad 500039, India. ✉e-mail: gintautas.vainorius@imba.oeaw.ac.at; ulrich.elling@imba.oeaw.ac.at

different cell types to functioning neurons[24–28]. Among those, Ascl1 and Ngn2 (encoded by the *Neurog2* gene) have emerged as preferred tools for in vitro neuronal reprogramming[24]. Both are master regulators of neuronal fate in the developing central nervous system. Expression of Ascl1 in ventral telencephalon progenitor cells generates inhibitory GABAergic interneurons, whereas Ngn2 directs dorsally situated progenitors toward excitatory neurons with glutamatergic identity. Although, Ascl1 and Ngn2 give rise to different subtypes in the telencephalon, expression of Ngn2 in the ventral telencephalon can rescue Ascl1 null mice[29]. Interestingly, while expression of Ascl1 and Ngn2 in astrocytes can recapitulate developmental neuronal subtype specification to GABAergic or glutamatergic neurons, respectively, expression of Ascl1 in mouse embryonic fibroblasts (MEF), as well as expression of Ascl1 or Ngn2 in mouse embryonic stem cells (ESC), lead to primarily glutamatergic neurons[27,28,30]. Thus, the subtype specification depends on the initial cell type. It is thus vital to understand how ectopically expressed TFs interact with the initial cellular context to define the outcome of conversions.

The ability to transition toward the same states using different transcription factors gives an opportunity to directly contrast differentiation mechanisms. In an elegant study, Aydin et al. showed that Ascl1 and Ngn2 in ESC prefer binding distinct E-box motifs[6]. This, in turn, initiates different accessibility and gene expression patterns, influencing downstream TFs, while still resulting in the formation of glutamatergic induced neurons (iN) in both cases[6,27,28,30]. However, downstream mechanistic differences, such as the downregulation of the initial pluripotency network, are still not understood. Furthermore, even though Ascl1 and Ngn2 possess 'pioneering' TF properties, their binding and expression can be influenced by cellular context, e.g., Ascl1 binds and induces muscle lineage when expressed in MEF, but myoblast induction was not reported in ESC[6,16,19].

In this work, we study the mechanistic differences in how the expression of Ascl1 or Ngn2 transitions cells between two identical states. We observe different side lineages forming in parallel to neurons due to differential binding and different strategies for exiting pluripotency upon Ascl1 and Ngn2 induction: Ascl1 rapidly shuts down the pluripotency network and arrests the cell cycle to install neuronal or trophoblast states, while Ngn2 retains *Sox2* expression to produce neuron stem cells (NSC). CRISPR/Cas9 forward genetic screens revealed that genes involved in pluripotency regulation and cell cycle control are affecting neuronal reprogramming by Ascl1, but not Ngn2. Our results highlight different mechanistic pathways to iN employed by these bHLH TFs.

Results

Ascl1 and Ngn2 convert ESC to iN but generate different side lineages

Ectopic expression of Ascl1 or Ngn2 in mouse embryonic stem cells (ESCs) is sufficient to induce terminal differentiation into neurons[28]. Yet, the differences in transition mechanism toward neurons as well as possible side populations are not well characterized[6,31]. To examine this cell type conversion in detail, we generated clonal ESC cell lines expressing rtTA and TetO-Ascl1 or TetO-Ngn2[27,28] (Supplementary Fig. 1a). After doxycycline (Dox) addition, ESCs are rapidly converted to induced neurons: Ascl1 and Ngn2 produce cells expressing the neuronal marker TUBB3 and displaying neuronal morphology from day 3 and day 2 onward, respectively (Supplementary Fig. 1a). To report neuronal fate in these cell lines we endogenously tagged the pan-neuronal marker gene Mapt on its C-terminus with the fluorescent protein Venus[27] (Supplementary Fig. 1b) and performed time-resolved bulk RNAseq upon Dox-induction. Cells were sorted into Venus-positive neurons and Venus-negative cell populations (Fig. 1a) from day 3 onward. As reported before[28,30], both Ascl1 and Ngn2 give rise to similar iN cell identities (Fig. 1b; Supplementary Fig. 1c–e). Thus, initial ESC and terminal iN states are very similar between Ascl1 and Ngn2-induced conversions (Supplementary Fig. 1b bottom). This is in line with previous observations that transcriptomes converge to drive iN formation despite differences in the initial transcriptional response[6] and follow an overall similar trajectory between ESC and iN in the PCA analysis (Fig. 1c, d; Supplementary Fig. 1g).

To investigate cells that fail to make iN in more detail, we focused on the Mapt-Venus-negative cells, which could represent incomplete or alternative differentiation outcomes. Mapt-Venus-negative cells generated by Ascl1 cluster closer to the initial ESC populations in PCA plots than Ngn2-induced Mapt-negative cells (Fig. 1c). However, this is not due to retaining a population of undifferentiated ESC as only a marginal number of cells express ESC marker NANOG in terminal population (Supplementary Fig. 1g), and NANOG and OCT4 are not expressed in Mapt-Venus population (Supplementary Fig. 1h). To identify the alternative type of cells generated by Ascl1, we used PanglaoDB[32] using genes differentially expressed between Venus-negative and positive populations at day 6 (Fig. 1e, f; Supplementary Fig. 2a). Interestingly, Ascl1 produce cells expressing trophoblast markers such as *Hand1*, *Cdx2*, *Tpbpa*, *Krt8* (Fig. 1e, g; Supplementary Fig. 2b–g) with mesenchymal morphology, which were not present in Ngn2-induced cultures (Fig. 1g, Supplementary Fig. 2b, c, f). We termed these cells induced-Trophoblast-like-cells (iT). Interestingly, many of the *Krt8*, *Cdx2* positive iTs appeared binucleated, which could be a result of multinucleation similar to trophoblast lineage development in vivo (Supplementary Fig. 2e)[33]. The induction of iT could be due to Ascl1 mimicking the bHLH transcription factor Ascl2, a driver for the trophoblast lineage. Both Ascl1 and Ascl2 are evolutionary close and share near identical DNA binding domains and bind similar E-box motifs (Supplementary Fig. 2h–k)[33–37]. Lastly, to exclude clonal effects of the cell line used, we repeated these experiments in the background of an alternative mouse ESC line, E14. We introduced rtTA via a piggybac transposon vector and expressed Ascl1 from a Dox-inducible viral vector[27] and generated 24 single-cell derived clones. All the clones showed the formation of both iN and iT, suggesting that the formation of iTs is a reproducible side product of ectopic Ascl1 expression in mouse ESCs (Supplementary Fig. 3).

In contrast to Ascl1 induction, Ngn2 induces Mapt-negative cells expressing NSC markers, such as *Sox2*, *Pax3*, *Pax6*, *Nes* (Fig. 1f, h; Supplementary Fig. 2a, 4a, b), as previously described[21,38]. Furthermore, Ngn2 reprogramming could be locked in the NSC-like state (iNSC) in the presence of FGF2 and EGF and is dependent on the Notch pathway[39,40] (Supplementary Fig. 4c, d). In contrast, we did not observe NSC markers upregulated during Ascl1-induced differentiation (Fig. 1h, Supplementary Fig. 4b, c). To see if Ngn2 can use iNSC state as proliferative intermediate, we differentiated cells in the presence or absence of cytosine β-D-arabinofuranoside (AraC) from day 4 post-induction to inactivate dividing cells (Supplementary Fig. 4e). Indeed, addition of AraC drastically reduces Ngn2-produced iNs, while Ascl1 was insensitive to AraC treatment, suggesting that no continuously proliferative intermediate is present during Ascl1-induced iN reprogramming (Supplementary Fig. 4e). In summary, despite Ascl1 and Ngn2 converting mESC to similar iN subtypes, Ascl1 and Ngn2 produce distinct additional alternative cell lineages, suggesting that despite identical initial and terminal populations, differences exist that we sought to understand further (Supplementary Fig. 1b bottom).

Ascl1 and Ngn2 initiate paths with different transcriptional programs

To get a better understanding of different transcriptional response invoked by Ascl1 and Ngn2, we performed bulk RNAseq and ChiPseq on day 1 (Fig. 1a). Both Ascl1 and Ngn2 induce general neuronal markers such as *Tubb3*, *Map2* and *Onecut2* and downregulate general pluripotency markers like *Nanog*, *Klf4* (Fig. 1i). Furthermore, Ascl1-strongly induces downstream targets *Tfap2b*, *Lmx1b*, while

https://doi.org/10.1038/s41467-023-40803-y

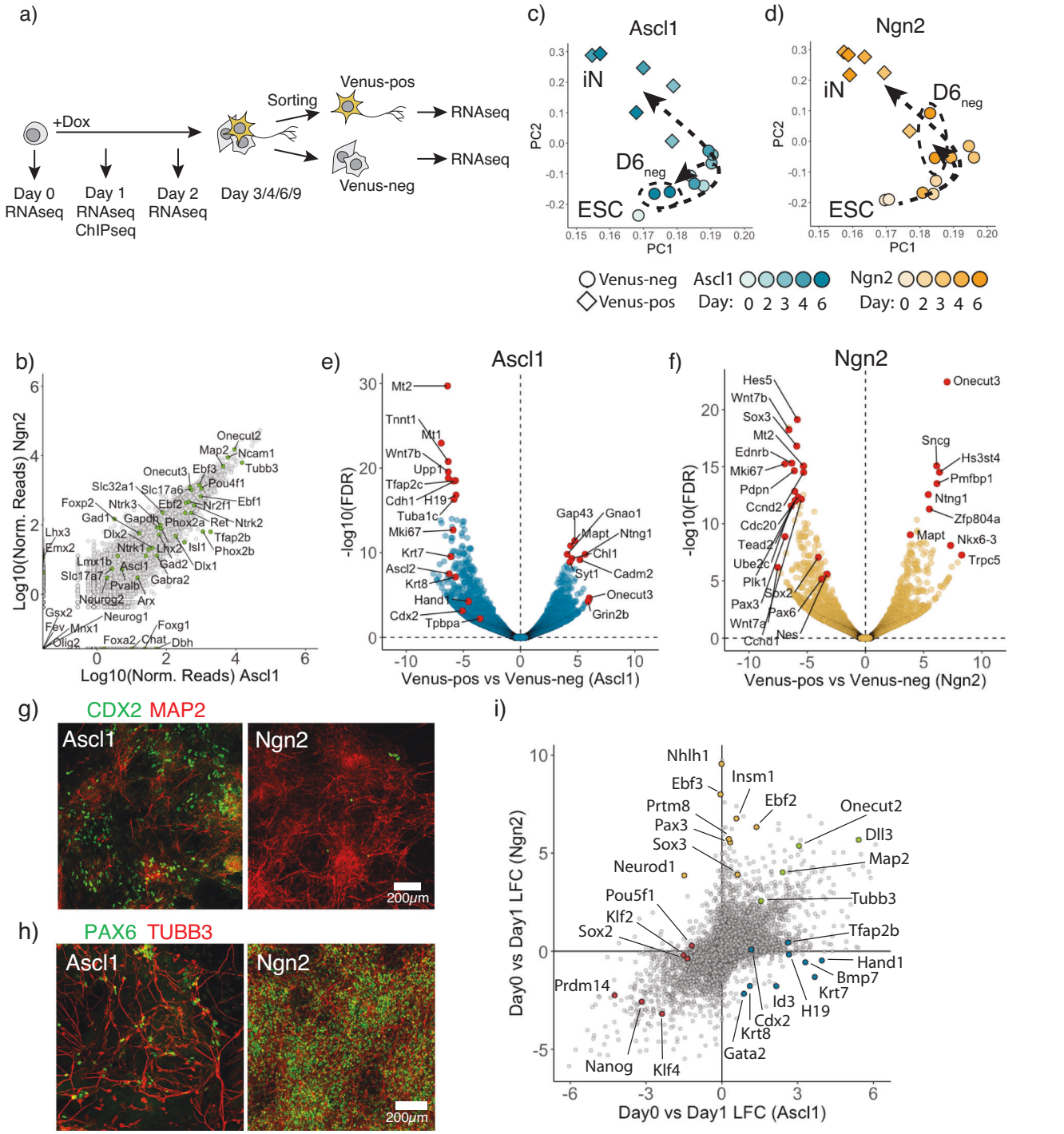


Fig. 1 | Ascl1 and Ngn2 induce different alternative lineages. a Schematic overview of the experimental design. **b** Scatter plot comparing gene expression at Day 6 between Ascl1 and Ngn2 Venus-positive cells with various neuronal subtype specific markers indicated in green. **c**, **d** Principal component analysis of time-resolved bulk RNAseq after Ascl1 (**c**) or Ngn2 (**d**). Each data point corresponds to the single time point replicate. Color intensity shows day post-induction. Shape corresponds the Mapt-Venus reporter upregulation. Arrows show the trajectory cells take after the Ascl1 (**c**) or Ngn2 induction (**d**). **e**, **f** Vulcano plot comparing gene expression between Venus-positive and negative populations at day 6 post-induction of Ascl1 (**e**) or Ngn2 (**f**). Red circles denote top significantly upregulated or downregulated genes as well as example genes marking in trophoblast (**e**) or NSC lineages (**f**). **g** Representative immunostained cells for a trophoblast marker CDX2 and a neuronal marker Map2 at day 6 post-induction of Ascl1 or Ngn2. Trophoblast markers were expressed only after Ascl1 induction, but not Ngn2. **h** Representative immunostained cells for an NSC marker PAX6 and a neuronal marker TUBB3 at day 6 post-induction of Ascl1 or Ngn2. NSC markers were expressed only after Ngn2 expression, but not Ascl1. **i** Scatter plot comparing gene expression changes between Ascl1 and Ngn2 at day 1 post-induction. Highlighted circles are example genes that are neuronal markers expressed in both (green), trophoblast Ascl1 specific markers (blue), NSC Ngn2 specific markers (yellow), pluripotency related genes (red). Source data are provided as a Source Data file.

Ngn2 strongly upregulates *Neurod1*, *Nhlh1* (Fig. 1i). In addition, Ascl1 upregulate Trophoblast lineage markers, e.g., *Krt7/8*, *Hand1*, while Ngn2 upregulates expression of NSC related genes, like *Pax3* and *Sox3* (Fig. 1i). Interestingly we observed that early in reprogramming cells are positive for both—neuronal and trophoblast markers (Supplementary Fig. 5a, b). In addition, we reanalyzed available scRNAseq data[6] for Day 2 of ESC to iN conversion by Ascl1 and Ngn2 and could also observe cells positive for both neuronal and trophoblast markers (Supplementary Fig. 5c, d). This suggests that Ascl1 can induce both lineages simultaneously, which later are resolved into iN or iT cells (Supplementary Fig. 5b).

As reported by Aydin and colleagues[6], Ascl1 and Ngn2 show different preferences for E-box motives, which in turn result in the activation of different subset of genes. Indeed, we confirmed Ascl1 and Ngn2 differential binding in ESC (Supplementary Fig. 6a). Furthermore, we see that Ascl1 and Ngn2 target genes differentially expressed between them as well as genes involved in the different alternative lineages, e.g., *Hand1*, *Cdx2*, *Krt8* or *Neurod1*, *Pax3*, respectively (Supplementary Fig. 6b). Interestingly, Ascl1 binds trophoblast related genes also in MEF, although without iT induction (Supplementary Fig. 7a, b). However, in addition to iN induction, Ascl1 overexpression in MEF leads to Ascl1 binding to the skeletal muscle genes and the induction of myocytes[7,16]. Indeed, in ESC Ascl1 also strongly bind to skeletal muscle lineage-related genes, e.g., *Myod1*, *Myog*, *Myf3*, *Tnnt2* (Supplementary Fig. 7c). However, we did not observe upregulation of these genes (Supplementary Fig. 7d). Thus, it is tempting to speculate that cellular context, such as cell type-specific histone modifications or transcription factors, affects the choice of the alternative lineage induced. Thus, in addition to the induction of the neuronal transcriptional program, additional genes are bound and transcribed, leading to the formation of alternative lineages. This "off target" transcriptional program depends on overexpression of transgene, e.g., Ascl1 or Ngn2, as well as the cellular context, e.g., ESC or MEF.

To investigate the general principles of the early transcriptional response in more depth, we used the STRING database to examine the gene regulatory network (GRN) of Ascl1 and Ngn2-upregulated differentially expressed genes (DEGs) one day after induction (Supplementary Fig. 8a, b). Both Ascl1 and Ngn2 DEGs form networks containing three distinct gene groups. Interestingly, Ascl1 and Ngn2 GRNs share groups containing genes involved in RNA and sterol metabolism, which can be attributed to the metabolic shift during conversion[41,42]. Furthermore, Ngn2 forms a more interconnected network than Ascl1, suggesting that Ngn2 invokes a more coherent transcriptional response than Ascl1 (Supplementary Fig. 8c).

To analyze the differences in more detail, we looked at the centralmost connected nodes of both GRNs (Supplementary Fig. 8d, e). In contrast to Ascl1, Ngn2 induce genes driving neuronal differentiation *Neurog2*, *Neurod1*, *Neurog1*, *Lhx2/3*, *Otx2*, as well as genes involved in neural stem cell differentiation: *Notch1*, *Hes5*, *Pax3*, centering GRN around them (Supplementary Fig. 8e–g). Interestingly, Ngn2, Neurod1 and Ngn2 are known to be able to convert ESC to iN[43]. Thus, such a positive feedback loop together with the induction of strong lineage drivers and highly interconnected GRN can allow a more robust iN conversion and faster independence from the initial induction of the cassette. To test this hypothesis, we induced cells for 2, 4, or 6 days (Supplementary Fig. 8h). Indeed, we observe that efficient iN conversion with Ascl1 relies on the sustained expression of Ascl1, while Ngn2 efficiently induces conversion already after 2 days of induction (Supplementary Fig. 8h).

In summary, Ascl1 and Ngn2 bind and invoke different transcriptional profiles and subsequent mechanistic differences of the ESC to iN conversion. Where Ascl1 induction relies on the sustained expression of Ascl1, Ngn2 induces an overall more coherent network allowing efficient and fast reprogramming as well as induction of a proliferative intermediate.

Loss-of-function CRISPR/Cas9 screen to identify genetic dependencies for ESC to iN conversion

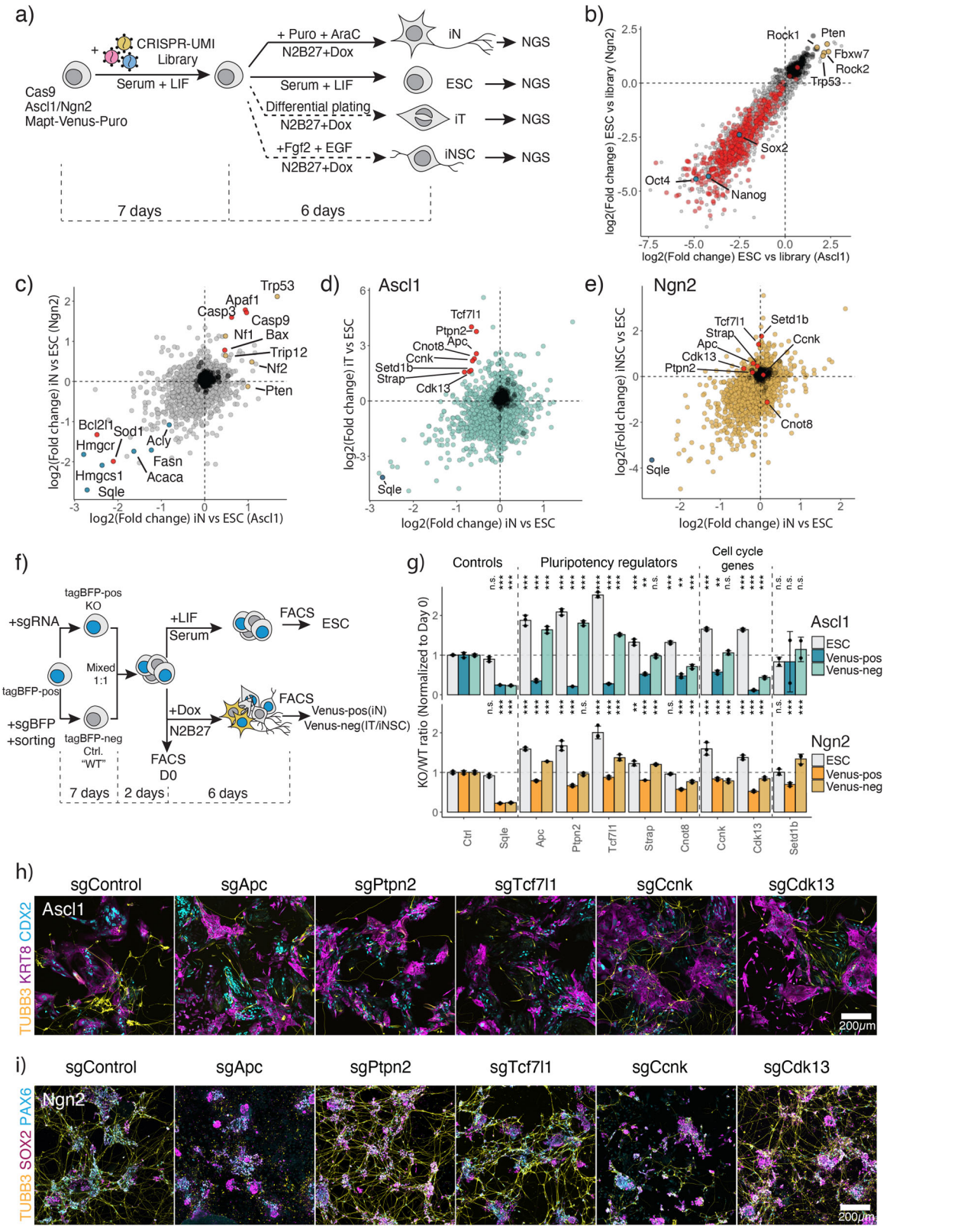
To establish a mechanistic handle on the underlying differences between Ascl1- and Ngn2-induced ESC reprogramming, we performed a CRISPR/Cas9 loss-of-function screen. We aimed to investigate two aspects of this conversion: first, the genetic dependencies of Ascl1 and Ngn2-induced neuronal conversion, and second, to elucidate the genes involved in alternative state formation (Fig. 2a). For this, ESC were infected with retroviral CRISPR-UMI sgRNA library containing ~27,000 guides targeting 6630 genes (4 guides per gene) and 108 non-targeting guides[44]. We then differentiated ESC for 6 days and enriched for iN using AraC and puromycin treatment (Fig. 2a, Supplementary Fig. 1b, 9a). In addition, we assessed dependencies for iT and iNSC differentiation (Fig. 2a, see "Methods"). For an initial assessment of the screening results, we compared the depletion of known essentials in ESC versus the Library (Fig. 2b). We indeed observed a strong depletion in common essential genes as defined by Hart et al.[45]. Furthermore, genes at the core pluripotency network (*Nanog*, *Sox2*, *Pou5f1*) are among the most essential genes, while tumor suppressors including *Trp53*, *Fbxw7*, *Rock1* are strongly enriching in ESC (Fig. 2b), showing that our library is effective at gene targeting and revealing gene knockout phenotypes.

To identify common genetic dependencies of iN formation, we compared guide abundance in the ESC versus iN at D6 (Fig. 2c). Guides showing the strongest depletion are targeting, for instance, fatty acid metabolism and sterol biosynthesis, such as *Fasn*, *Hmgcr*, *Hmgcs1*, *Sqle*, an essential component for neuronal metabolism (Fig. 2c)[41,42]. Furthermore, during iN differentiation, cells undergo high levels of stress due to lipid peroxidation and can commit cell death via apoptosis or ferroptosis[46,47]. Hence, we find proapoptotic genes, *Casp3/9* and *Bax* enriching, while antiapoptotic genes such as *Bcl2l1* and *Sod1* are depleting from the population (Fig. 2c; Supplementary Fig. 9b). In addition, knockout of tumor suppressors, e.g., *Trp53*, *Nf1*, *Nf2*, allows for higher iN conversion. Thus, this screen yields genetic dependencies in ESC to iN conversion at high resolution.

To uncover differential genetic dependencies between iN and alternative states, relative sgRNA abundance was correlated with a particular focus on genes depleted in iN and enriched in the alternative states. We focused on differential dependencies between the Ascl1-induced iN and iT (Fig. 2d). We picked genes for validation based on the following criteria: (1) genes showing little effects in ESC ($-2 < LFC < 2$); (2) genes showing enrichment in iT ($LFC > 1.5$); (3) genes depleted in iN ($LFC < -0.5$). We noticed that genes important in the regulation of pluripotency network and differentiation of ESC: *Tcf7l1*, *Ptpn2*, *Apc*, *Strap*, *Cnot8*[48–54], as well as cell cycle genes *Ccnk* and *Cdk13*[55,56] are required for Ascl1-induced iN conversion while inhibiting iT formation. In contrast, knockout of these hits had little to no influence on Ngn2-driven conversion to neurons (Fig. 2d, e).

To validate the findings, we performed a competition assay. To implement an internal control, we introduced a constitutive tagBFP vector in our cell line and derived a clone with a stable tagBFP expression. The cells were then infected with the sgRNA that showed the strongest depletion in the screen. Separately, we generated an isogenic control population, by targeting tagBFP with a control guide against tagBFP and sorting for $tagBFP^{neg}$ cells. Subsequently, knockout and control cells were mixed in a 1:1 ratio and maintained for an additional passage to adapt them to the same conditions, before plating cells for induction (Fig. 2f). After 6 days of induction, we assessed iN Mapt-Venus-positive population, as well as Venus-negative population, corresponding to the formation of iT/iNSC populations (Fig. 2f, g). In addition, we assessed the fitness of the knockouts in the ESC cells while growing ESC for additional 6 days (Fig. 2f). As anticipated, *Sqle* was essential in iN, iT or iNSC populations, while not in ESC cells, mirroring the screen results (Fig. 2c–e, g). Similarly, knockout of pluripotency-related genes *Apc*, *Tcf7l1* and *Ptpn2* abolished the

https://doi.org/10.1038/s41467-023-40803-y



formation of iNs upon induction with Ascl1, while allowing the formation of iT cells (Fig. 2g). We further confirmed the phenotype via immunostaining at day 6 post-induction (Fig. 2h, i). Knockout of the hits resulted in failure to generate TUBB3 positive neurons by Ascl1 while still generating CDX2+ and KRT8+ cells. Interestingly, some cells expressing neuron-specific TUBB3 lack neuronal morphology, indicating activation of neuronal genes but failure to establish a final iN state (Supplementary Fig. 9c). Likewise, targeting of *Ccnk* or *Cdk13* strongly reduced the number of iNs upon Ascl1 induction. In contrast to Ascl1, Ngn2 iN conversion is not impaired upon loss of *Tcf7l1*, *Ptpn2*, and *Cdk13* and Ngn2 can still generate both neurons and neural stem cells, marked by SOX2 and PAX6 (Fig. 2g–i). However, *Apc* and *Ccnk* knockouts did affect the formation of iN and iNSC induced by Ngn2. Off note, we additionally validated multiple other genes showing

Fig. 2 | CRISPR-Cas9 forward genetic screen to identify genetic dependencies of forward differentiation by Ascl1 or Ngn2. a CRISPR-Cas9 screen experimental outline. **b** Comparative analysis of the gene knockout effects between Ascl1 and Ngn2 transgenes carrying ESC. Difference in guide abundance was calculated between uninduced ESC at Day 13 post library infection (**a**) versus library plasmid pool. Dots represent genes; axis shows depletion in LFC in each cell line. Red represents core essential genes as defined by Hart et al.[45], blue—core pluripotency genes, yellow—tumor suppressors. **c** Comparative analysis of gene knockout effects of Ascl1 or Ngn2-induced iN versus ESC. Red dots represent apoptosis related genes, yellow—tumor suppressor genes, blue—cholesterol biosynthesis genes. **d** Comparative analysis of gene knockout effects of Ascl1-induced iN or iT versus ESC. Red dots represent genes chosen for validation. Sqle is a positive control for a strong depletion. **e** Comparative analysis of gene knockout effects of Ngn2-induced iN or iNSC versus ESC. Red dots represent genes chosen for validation in (**d**). Sqle is a positive control for a strong depletion. **f** Experimental outline of hits validation from (**d**). **g** FACS-based validation of the hits. Bar represents the normalized ratio to the initial mixture ratio at day 0. $N = 3$ independent biological replicates. Bar plot shows mean ± SD. p-values, indicated above, were determined by one-way ANOVA followed by Dunnett's multiple comparison test (two-sided) using Ctrl ratio as a control. "n.s." not significant, "*" $p < 0.05$, "**" $p < 0.01$, "***" $p < 0.001$. **h**, **i** Immunostaining of the knockout cells for neuronal marker Tubb3 and alternative lineage markers KRT8/CDX2 for Ascl1-induced iT (**h**), and SOX2/PAX6 for Ngn2-induced iNSC (**i**). Source data are provided as a Source Data file.

differential dependencies between Ascl1 and Ngn2 induction (Supplementary Fig. 9d–h). Taken together, our parallel CRISPR/Cas9 loss-of-function screens unveiled multiple common and differential dependencies between Ascl1- and Ngn2-induced directed differentiation to iNs.

Rapid downregulation of pluripotency network upon Ascl1 induction

As the screen identified genes involved in maintaining the pluripotency network as essential for the Ascl1-induced iN formation, we investigated how Ascl1 and Ngn2 disassemble the pluripotency network in more depth. For this, we used ingenuity pathway analysis using all differential expressed genes one day post Ascl1- or Ngn2-induction (Supplementary Fig. 10a–d, Fig. 1i). As before, ingenuity canonical pathway enrichment analysis showed that both Ascl1 and Ngn2 induce pathways related to cholesterol biosynthesis (Supplementary Fig. 10b, d). Interestingly, the Ascl1 transcriptional response is centered around the downregulation of the ESC pluripotency network and self-renewal (Supplementary Fig. 10a, b).

We then divided DEGs into Ascl1 or Ngn2 specific, or common and performed KEGG pathway enrichment analysis (Fig. 3a, b; Supplementary Fig. 10e). Interestingly, Ascl1 upregulates genes involved in cellular senescence and cell cycle exit, such as *Cdkn1a, Cebpa*, *Cebpb* (Supplementary Fig. 8d, g)[57,58], as well as downregulates more genes involved in pluripotency, e.g., *Pou5f1* (encoding for protein OCT4), *Klf2*, *Sox2*, *Lef1*, *Lefty2*, compared to Ngn2 (Fig. 3b, Fig. 1i, Supplementary Fig. 10e). Thus, we next looked at the dynamics of the pluripotency network (PPN) shutdown upon Ascl1 and Ngn2 induction. For this, we focused on the three core pluripotency genes: *Pou5f1* (Oct4), *Sox2*, *Nanog* (Fig. 3c, Fig. 1i). *Nanog* is downregulated at similar kinetics between Ascl1 and Ngn2 (Fig. 3c, Fig. 1i, Supplementary Fig. 11a, b). In contrast, *Sox2* expression is retained after Ngn2 induction as an NSC gene regulatory network is established, while Ascl1-induced cells lose expression of *Sox2* (Fig. 3c, Fig. 1i, Supplementary Fig. 11a, b). Similarly, Ascl1 induction leads to a rapid loss of Oct4 expression, while Ngn2 induction leads to gradual downregulation of Oct4 (Fig. 3c–e, Fig. 1i). Furthermore, loss of Oct4 corresponds with the upregulation of trophoblast marker KRT8 (Fig. 3f).

To test the functional relevance of the loss of Oct4, we constitutively overexpressed Oct4 and generated multiple clonal mESC cell lines (Fig. 3g, h). While overexpression of Oct4 together with Ngn2 leads to a reduction of the iN population, co-expression of Oct4 and Ascl1 increases the efficiency of iN formation (Fig. 3g, h). In turn, iT formation is hampered in the presence of Oct4 (Fig. 3h). During development, Oct4 inhibits differentiation of the trophectoderm lineage, and rapid loss of Oct4 is associated with the upregulation of the trophectoderm markers[59,60]. Thus, rapid loss of Oct4 generates permissive conditions for iT lineage formation, that compete with the formation of iN.

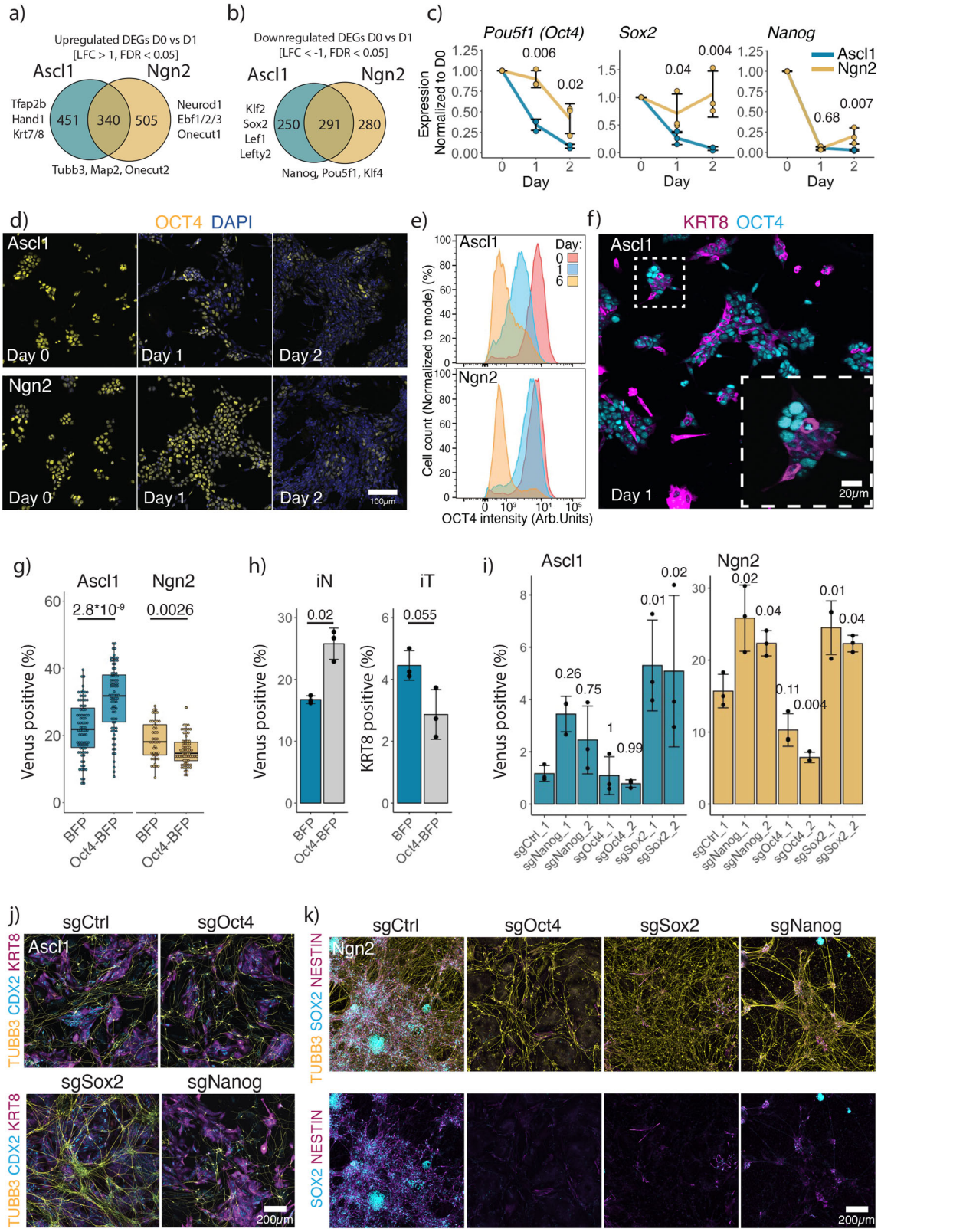
Given the kinetic differences in the downregulation of PPN after Ascl1 or Ngn2 induction, we tested if enforced disruption of the PPN together with differentiation would affect the ESC to iN conversion. As *Pou5f1* (Oct4), *Sox2* and *Nanog* are essential for ESC (Fig. 2b), we infected cells with the guides targeting *Pou5f1* (Oct4), *Sox2* and *Nanog* 2 days before the induction of Ascl1 or Ngn2 (Fig. 3i–k). Interestingly, Ascl1 induction leads to the formation of both iN and iT in the knockout of all pluripotency factors (Fig. 3j). We see little to no effect after targeting Oct4, as Oct4 is rapidly lost upon Ascl1 induction (Fig. 3i). However, knockout of Nanog or *Sox2* favors iN formation, suggesting that additional disruption of the PPN supports installment of the iN state (Fig. 3i, j). In turn, knockout of *Sox2* abolished the formation of iNSC by Ngn2, showing that *Sox2* is repurposed from the PPN to the NSC gene regulatory network (Fig. 3i, k). Similarly, disruption of either *Nanog* or *Pou5f1* (Oct4) also resulted in loss of iNSC (Fig. 3k). Instead, we observed formation of primitive endoderm and trophoblast upon loss of *Nanog* or *Pou5f1* (Oct4), respectively (Fig. 3k, Supplementary Fig. 11c, d), in alignment with the outcome of the loss of these factors during the development[59–61].

This data, together with differential dependencies identified in the CRISPR screen, shows that Ascl1 and Ngn2 induction leads to pronounced functional differences in exiting the pluripotency state of ESC. Ascl1 induction leads to an efficient shutting down of PPN and later induction of iN or iT gene network, while Ngn2 does not fully downregulate the PPN and instead overlays both networks, repurposing genes for the induction of NSCs.

Tcf7l1 is required for cell cycle exit to generate Ascl1 iNs

We next focused on Ascl1-induced PPN shutdown. In Ascl1-driven MEF to iN transdifferentiation, Myt1l facilitates iN formation by repressing initial fibroblast GRN as well as alternative myoblast lineage[16,62]. Given the fast downregulation of PPN, we tested if Myt1l is similarly required for ESC to iN conversion. However, our screen did not reveal that *Myt1l*, or its paralogs *Myt1* and *St18*, are required for ESC to iN conversion (Supplementary Fig. 12a) as well as validation experiments using additional sgRNAs (Supplementary Fig. 12b). In the CRISPR screen we identified *Tcf7l1* (T cell factor/lymphoid enhancer factor, also known as Tcf3), a repressor of PPN, as essential for iN formation by Ascl1. We confirmed the *Tcf7l1* knockout phenotype in a separate E14 background and ruled out that the loss of Tcf7l1 affected the expression of the Ascl1 transgene (Supplementary Fig. 12c–e). Tcf7l1 is poised on the multiple promoters of PPN-related genes and rapidly represses their expression upon ESC differentiation[48,63,64], and the absence of Tcf7l1 stabilizes the ESC state[64]. Thus, we wanted to test if Tcf7l1 is required before or after the onset of cell type conversion, as well as exclude secondary effects due to the stabilization of the PPN before the conversion. For this, we N-terminally tagged Tcf7l1 with an Auxin inducible degron, infected cells with lentiviral vector carrying osTIR1 (F-box E3-ubiquitin ligase, derived from Oryza sativa) and generated single cell-derived clones (Supplementary Fig. 12f)[65]. We then depleted Tcf7l1 before or at the onset of conversion by the addition of Auxin (Supplementary Fig. 12g). While Tcf7l1 depletion before induction had only a mild effect on the formation of iN, degrading Tcf7l1 on the onset of Ascl1 induction severely reduced the number of neurons produced by

https://doi.org/10.1038/s41467-023-40803-y



Ascl1 (Supplementary Fig. 12g). This indicates that Tcf7l1 acts after the induction rather than stabilizing the PPN.

To understand the role of Tcf7l1 in the Ascl1-induced ESC to iN-directed differentiation, we generated single cell-derived clones with a homozygous *Tcf7l1* knockout and performed bulk RNAseq at day 1 post-induction (Supplementary Fig. 13a). As Tcf7l1 acts as a pluripotency network repressor during ESC differentiation, we first looked at the pluripotency shutdown in the *Tcf7l1* knockout cells. Surprisingly, the PPN was still downregulated after Ascl1 induction in the absence of Tcf7l1 (Fig. 4a; Supplementary Fig. 13a). However, we observed a group of genes that failed to be expressed in *ΔTcf7l1* clones (Supplementary Fig. 13a). Interestingly, *Cdkn1c* is highly upregulated after Ascl1 induction in comparison to *ΔTcf7l1* (Supplementary Fig. 13a), and its expression is specific for Ascl1 induction and peaks at day 3 of ESC to iN

https://doi.org/10.1038/s41467-023-40803-y

Fig. 3 | Dynamics of pluripotency network shutdown. a, **b** Number of Ascl1, Ngn2 specific or common upregulated (**a**) and downregulated (**b**) genes from Fig. 1i. **c** qPCR data of the expression of the core pluripotency genes (normalized to the expression of Actin and day 0). Lines are drawn through the mean of $n = 3$ biologically independent samples; error bars indicate ± SD. Above, p-values of the two-sided Welch two-sample t-test comparing PPN genes expressed between Ascl1 and Ngn2 at the given time point. **d** Representative immunostainings for the OCT4 dynamics after Ascl1 or Ngn2 induction. **e** Quantification of OCT4 expression using intracellular immunostaining followed by FACS. **f** Representative immunostaining for the pluripotency marker OCT4 and trophoblast marker KRT8 at day 1 after Ascl1 or Ngn2 induction. **g** Efficiency of iN formation in the presence of OCT4 overexpression. Each dot represents an individual ESC clone containing an overexpression construct. Efficiency is measured by the percentage of Mapt-Venus population. Boxplots indicate 25th and 75th percentiles as bounds of the box with the median center line; whiskers indicate minima/maxima of a 1.5x distance of the IQR from the 25th and 75th percentiles. p-value of the two-sided Welch two-sample t-test indicated above. **h** Efficiency of iN formation, measured by the percentage of Mapt-Venus expressing cells, and iT formation, percentage of cells immunostained for KRT8 of polyclonal population overexpressing Oct4-BFP. Bar plot shows mean of $n = 3$ independent biological replicates with ± SD; the p-values were calculated using the two-sided Welch two-sample t-test. **i** Efficiency of iN formation upon acute knockout of core pluripotency. Efficiency is measured by the percentage of Mapt-Venus population. The bar plot shows mean of $n = 3$ biologically independent samples with ± SD. p-values, indicated above, were calculated using one-way ANOVA followed by Dunnett's multiple comparison test (two-sided) using sgCtrl as a control. **j** Representative immunostainings of (**i**) for iN and iT induced by Ascl1. **j** Representative immunostainings of (**i**) for iN and iNSC induced by Ngn2. Source data are provided as a Source Data file.

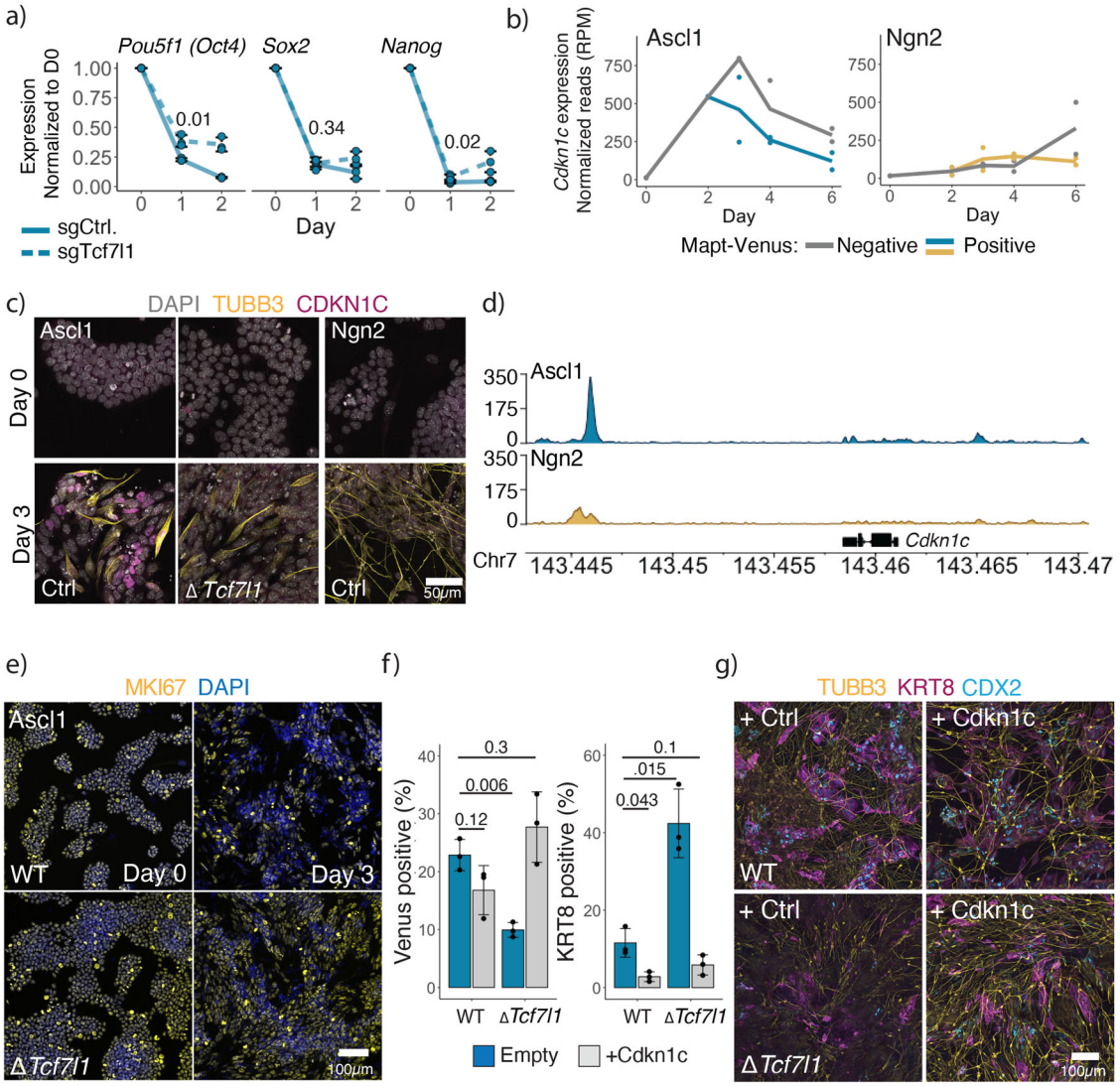


Fig. 4 | Overexpression of Cdkn1c rescues Tcf7l1 knockout phenotype. a qPCR of core pluripotency gene expression after induction in WT and *Tcf7l1* KO cells. Expression normalized to Actin and day 0. Lines are drawn through the mean of $n = 3$ biologically independent samples; error bars indicate ± SD. p-values of the two-sided Welch two-sample t-test comparing sgTcf1l1 vs sgControl at day 1 indicated above. **b** *Cdkn1c* expression during ESC to iN conversion (Fig. 1a). **c** Immunostainings of CDKN1C on day 0 and day 3 of the ESC to iN conversion by Ascl1 WT or *Tcf7l1* KO or Ngn2 expressing cells. **d** Binding of the Flag-Ascl1 or Flag-Ngn2 in the *Cdkn1c* locus at day 1 post-induction. Data showing combined reads of four replicates. **e** Immunostainings of proliferation marker MKI67 after induction of Ascl1 in WT or *Tcf7l1* KO ESC cells. **f** FACS data of Cdkn1c overexpression with Ascl1 in WT or *Tcf7l1* KO ESCs. The efficiency of iN formation is measured by the percentage of Mapt-Venus population, efficiency of iT formation is measured by the percentage of cells immunostained for KRT8. The bar plot shows mean of $n = 3$ independent biological replicates with ± SD. p-value of the two-sided Welch two-sample t-test indicated above. **g** Representative images of immunostained cells for neuronal TUBB3 and trophoblast CDX2/KRT8 markers at day 6 post-induction of WT or *Tcf7l1* KO ESCs with Ascl1 and Cdkn1c. Source data are provided as a Source Data file.

conversion (Fig. 4b, c; Supplementary Fig. 13b). In addition, Ascl1 shows strong binding near *Cdkn1c* locus in comparison to Ngn2 (Fig. 4d).

Ascl1-directed differentiation is cell cycle-dependent

Cdkn1c is a cyclin-dependent kinase inhibitor of the cip/kip family regulating cell cycle arrest in G1[66]. Cdkn1c is essential for embryonic development, and mice lacking Cdkn1c die perinatally with multiple developmental defects. Furthermore, Cdkn1c is important for the development of early placenta to initiate the endoreplication of trophoblasts[66]. Recently, Cdkn1c was also implicated in the suppression of pluripotency in mouse ESC[67]. We furthermore see that Ascl1-induced GRN central nodes contain multiple genes involved in cell cycle regulation, e.g., *Cdkn1a*, *Cdkn1b*, *Cebpa*, *Cebpb* (Supplementary Fig. 8d, g). Thus, we hypothesized that in *ΔTcf7l1* cells, Ascl1 is unable to arrest the cell cycle. Indeed, we observe prolonged expression of MKI67 as well as a higher percentage of dividing cells after induction of Ascl1 in the *ΔTcf7l1* cells (Fig. 4e, Supplementary Fig. 13c). To test if the expression of Cdkn1c is sufficient to induce iN formation in the absence of Tcf7l1, we coexpressed Cdkn1c together with Ascl1. Indeed, we see a partial rescue of iN formation as well as a decrease in iT generation (Fig. 4f, g). Furthermore, we see that expression of Cdkn1c generates a more mature neuronal and trophoblast phenotype also in WT populations by Day 6 (Fig. 4g). Of note, Cdkn1c is not essential for the formation of Ascl1-induced iNs, suggesting that a group of cell cycle regulators, rather than Cdkn1c alone, is responsible for the cell cycle arrest (Supplementary Fig. 13d).

Taken together, our data show that in Ascl1-induced ESC to iN conversion, cell cycle arrest is a roadblock after exiting the pluripotency state and that Cdkn1c is sufficient to overcome this roadblock. In contrast, this dependency is not observed for Ngn2-dependent iN formation, where the cell cycle is maintained and some ESCs transit to NSCs.

Discussion

Overexpressing proneural bHLH transcription factors Ascl1 or Ngn2 in mouse ESC show several differences in the conversion toward iN: (1) formation of different alternative lineages; (2) early transcriptional changes in downregulation of pluripotency network and upregulation of new transcripts; (3) different genetic dependencies. In addition to iN, Ascl1 induces trophoblast-like cells marked by Krt8 and Cdx2, while Ngn2 invokes an NSC-like state[38]. The induction of alternative states results from the binding and upregulation of alternative lineage specifying factors, e.g., *Hand1*, *Cdx2* in Ascl1 case and *Pax3* in Ngn2. To induce trophoblast lineage, Ascl1 in mESC acts similarly to Ascl2, a key trophoblast lineage driving the bHLH transcription factor. Interestingly, during MEF to iN transdifferentiation, Ascl1 also binds trophoblast lineage drivers in MEF, although without inducing them (Supplementary Fig. 7a, b). Instead, additionally to iN, Ascl1 expressed in MEF induces myocyte-like cells[16,19]. In turn, muscle lineage genes and other targets are bound in both ESC and MEF[16,19] (Supplementary Fig. 7c, d). Thus, besides the neuronal lineage, only those Ascl1-bound genes are induced, which are developmentally close to the initial cell lineage in which Ascl1 is expressed. Differences in epigenetic landscape, histone modifications around binding targets as well as interaction partners or cofactors present in the initial cell types could specify binding and induction of target genes molding the mechanism[6–9,12,14]. Thus, the iN conversion paradigm from different cell types using the same transcription factor provides a powerful experimental platform to study how transcription factors binding and induction of lineages can be affected by cellular context.

An important aspect of lineage conversion is the efficient inhibition of the initial cell type. Overcoming the initial state is a well-described bottleneck in reprogramming to iPS cells[68,69], and utilizing efficient initial lineage repression would require fewer cues to install a new cell identity. For example, miRNA-induced neuronal reprogramming of fibroblasts primarily acts via repression of the REST complex and non-neuronal factors, highlighting the importance of repression of the initial cell state[70]. Even though Ascl1 acts primarily as a transcriptional activator[6], we show that Ascl1 rapidly downregulates the PPN. *Pou5f1* (Oct4) and *Nanog* are lost within 24 h upon Ascl1 expression, and *Sox2* is largely downregulated after 48 h. Similarly, MEF gene regulatory network is rapidly disassembled upon Ascl1 expression[7,71]. However, the Ascl1-induced pluripotency repression mechanism is yet to be addressed.

We conducted comparative forward genetic screens to systematically elucidate the genetic dependencies of both Ascl1 and Ngn2-induced transitions. Interestingly, Ascl1 transition was more dependent on the factors modulating the PPN and cell cycle. We selected Tcf7l1, a well-described PPN repressor, as a representative of this gene group. Surprisingly, the PPN was still downregulated even after *Tcf7l1* knockout. However, we noticed that Ascl1 fails to induce Cdkn1c in the absence of Tcf7l1. This suggests that exit from the cell cycle might be a prerequisite for Ascl1-driven neuronal conversion. Indeed, co-expression of Cdkn1c with Ascl1 triggered cell cycle exit and rescued the loss of Tcf7l1. Taken together, the ability of Ascl1 to rapidly downregulate the initial cell state and effectively induce cell cycle arrest can be an inherent part of its strong transdifferentiation potential across large developmental distances[27,72]. In such a case, efficient downregulation of the initial network would also allow easier installment of the new cellular state even from the weaker GRN induced by Ascl1. However, multiple signals induced by Ascl1 in the absence of initial state GRN would also allow permissive conditions for installing alternative cell fates, e.g., iT in ESC and myocyte in MEF.

In contrast, Ngn2 induces a stronger and more interconnected and self-sustaining network than Ascl1 in ESC. Indeed, Ngn2 gives rise to iN cells faster and more efficiently (Supplementary Fig. 1a). Similarly, Ngn2 binds to and induces more targets than Ascl1 also in human fibroblasts, including proneural genes like *NEUROD1* and *NEUROD4*[73,74]. Yet, only Ascl1 alone is sufficient to convert MEFs to functional neurons, while Ngn2 requires the addition of small molecules to facilitate genome accessibility[74,75] or co-expression with additional transcriptional activators such as Brn3a[76,77]. This discrepancy could, at least in part, be explained by weaker repression of the initial state by Ngn2. For example, co-expression of transcriptional inhibitors, such as Myt1l, together with Ngn2 leads to successful MEF to iN conversion[62,76]. However, despite strong proneural GRN upregulation, Ngn2 induction leads to slower and incomplete repression of PPN after induction in ESC. This is indicative by the formation of trophoblast or primitive endoderm cells forming after and acute knockout of *Nanog* or Oct4, respectively, driven by the residual PPN[59–61]. Furthermore, Ngn2 retains *Sox2* expression as part of the NSC GRN, which acts as a proliferative intermediate. Thus, in ESCs, Ngn2 appears to overlay the PPN with neuron-specific genes and thereby generate a feed-forward loop to drive neuron formation with the appearance of NSC.

These observations indicate that Ascl1 and Ngn2 utilize different mechanisms to transition cells between identical states, namely from mESC to the same iN subtype. Specifically, Ascl1 dismantles one fate followed by the installment of another, while Ngn2 overlays the cell identities toward a new cell fate in a development-like process.

Methods

The experiments in this study are in compliance with relevant guidelines and ethical regulations.

Cell culture

Mouse ESCs (mESCs) were maintained in ESCM medium: DMEM (Sigma, D1152) supplemented with 15% Fetal bovine serum (FBS, Gibco, 10270106), 100 U/ml Penicillin, 100 µg/ml Streptomycin (Sigma, P0781), 1x Non-essential amino acids (Sigma, M7145), 4 mM L-Glutamine (Sigma, G7513), 1 mM Na-Pyruvate (Sigma, S8636), 0.1 mM

β-mercaptoethanol (Merck, 805740), 50 µg/ml ascorbic acid, 1000 U/ml LIF (ESGRO Millipore). Cells were trypsinized and replated every second day on gelatin-coated plates with the irradiated DR4 mouse embryonic fibroblasts (MEF) feeders, isolated from the E13.5 stage embryo of the DR4 mouse strain (RRID:I MSR_JAX:003208), plated a day before.

For iN induction, mESC were trypsinized, and 10^8 cells per well were plated on Matrigel (1 h 37 °C, 300 µL of 1000x diluted Matrigel (Szabo Scandic, BDL356231)) in a 24 well in the evening. The next day cells were washed twice with PBS and induced with N2B27 medium: DMEM/F12 (Gibco, 10565018), Neurobasal (Gibco, 21103049), N2 supplement (Gibco, 17502048), B27 supplement (Gibco, 17504044), 100 U/ml of Penicillin, 100 µg/ml Streptomycin (Sigma, P0781), 1x Non-essential amino acids (Sigma, M7145), 4 mM of L-Glutamine (Sigma, G7513), 1 mM of Na-Pyruvate (Sigma, S8636), 20 µg/mL Insulin (Roche, 15898200), 1 µg/ml doxycycline (dox). Time of induction was considered as day 0. From Day 2 onward, half of the medium was exchanged daily. For the induction of the iNSC state, cells were grown in the presence of 20 ng/ml FGF2 (PeproTech, 450-33) and 20 ng/ml EGF (PeproTech, 315-09). To enrich for non-dividing cells, 4 µM cytosine β-D-arabinofuranoside (AraC, Sigma-Aldrich) was added from day 4 onward. Cells were dissociated with trypsin and quantified using FACS BD LSR Fortessa (BD Biosciences). Subsequent gating of the data was done using Flowjo (BD Biosciences, v10.8.1).

Cell lines

Doxycycline-inducible TetO-Ascl1 and TetO-Ngn2 mouse ESC lines were established as described previously[7]. Cells were then infected with a lentiviral construct containing constitutive Cas9. and single cell-derived clones were tested for stable Cas9 expression and robust ESC to iN conversion. Plasmid containing V5-P2A-Venus-T2A-PuroTK-LoxP-Ef1a-Neo-LoxP tag flanked with 750 bp homology arms adjacent to *Mapt* terminus was co-electroporated with a CRISPR-Cas9 sgRNA targeting *Mapt* gene C-terminus (Supplementary Table 1) using Mouse Embryonic Stem Cell Nucleofector™ kit (Lonza) and selected with 0.5 mg/ml G418 (Gibco). Neomycin resistance was excised via transfection of the plasmid containing Cre and individual clones were derived, genotyped, and tested for the reporter activity via FACS and immunostaining.

Individual clones with *Tcf7l1* knockouts were derived by infecting cells with the lentiviral vector containing sgRNA targeting *Tcf7l1* (Supplementary Table 1), mutagenizing for 7 days and subsequent picking of the single cell-derived clones. The correct genotype was confirmed by Sanger sequencing of the endogenous *Tcf7l1* locus and western blot.

N-terminus Tcf7l1 tagging with AID degron was done by co-electroporating plasmid containing mCherry-V5-AID flanked by 1kbp homology arms and a plasmid with sgRNA targeting *Tcf7l1* start codon. Successful homozygous integration in single cell-derived clones was confirmed via genotyping and western blot. Cells were infected with lentiviral vector delivering pSFFV-osTIR-T2A-eBFP2 and single cell-derived clones were tested for degradation of Tcf7l1, retained BFP expression during differentiation and showed no hypomorphic effects on iN conversion without induction of degradation.

An alternative cell line for validations of the formation of alternative lineages and ChIP-seq experiments was derived by electroporating E14 mESC (RRID:CVCL_C320) with piggyBac vector containing CAG-rtTA-IRES-Hygro (Addgene plasmid number #102423) and a plasmid with piggyBac transposase. Cells were then infected with a lentiviral construct containing Tet-O-FLAG-Ascl1-T2A-Puro or Tet-O-FLAG-Ngn2-T2A-Puro (modified from Addgene plasmid #52047) and single cell-derived clones were subsequently tested for the resistance to puro after dox addition.

To confirm the *Tcf7l1* knockout phenotype, cells were derived in triplicates in the same way described in the ChIP experiment, but without clonal expansion. Cells were then infected with an U6-sgTcf7l1-PGK-Cas9-P2A-Blast lentiviral vector, selected with 10 µg/ml blasticidin (Invivogen) for 4 days, and mutagenized for a week before assessing the phenotype.

For Cdkn1c co-expression with Ascl1, cells were infected with Tet-O-Cdkn1c-T2A-Puro (modified from Addgene plasmid #52047) vector. After 3 days, cells were induced as described above. From day 1, Cdkn1c expression was selected with 1 µg/ml Puromycin (Invivogen).

For PPN perturbation experiments, sgRNAs against *Nanog*, *Pou5f1*, *Sox2* were cloned into spCas9-P2A-Puro vector. One day post-infection, cells were selected with 1 µg/ml Puromycin (Invivogen) for a day. A kill control without Puromycin resistance cassette was used to ensure complete selection. Cells were then split and induced with Dox, as described above.

SgRNA cloning and virus preparation

Individual sgRNAs (Supplementary Table 1) were cloned into the lentiviral plasmid containing spCas9-P2A-Blast using the single guide RNA Gecko cloning protocol (https://www.addgene.org/crispr/zhang). Lentiviral vectors were transfected using Polyethylenimine (PEI) into Lenti-X (Clontech, 632180) for virus production according to the supplier's recommendations. Virus containing supernatant was filtered through a 0.45 µm PES filter (VWR). For infection, 10^8 mESC were plated per gelatin-coated well in a 24-well plate. After recovery for at least 4 h, cells were infected with virus diluted 2x in fresh ESCM. The next day medium was exchanged and irradiated DR4 MEF feeders were plated on top of ESC. Successful integrations were selected with 10 µg/ml blasticidin for 4 days.

Immunofluorescence

Cells were grown and differentiated on the Matrigel-coated glass coverslips. Cultured cells were then gently washed with PBS, fixed for 15 min at room temperature with 4% paraformaldehyde (PFA) in PBS. Cells were washed twice with PBS and subsequently permeabilized and blocked with blocking solution (5% FBS, 0.1% Triton™ X-100 in PBS) for 1 h at room temperature. Then coverslips were incubated with primary antibody overnight at 4 °C. Used primary antibodies were mouse anti-TUBB3 (Sigma, T8660, 1:500), rabbit anti-TUBB3 (Biolegend/Covance, PRB-435P, 1:500), rabbit anti-MAP2 (Abcam, ab32454, 1:500), rat anti-SOX2 (Invitrogen (eBioscience), 14-9811-80, 1:500), rabbit anti-OCT4 (Abcam, ab19857, 1:500), rabbit anti-NANOG (Abcam, ab80892, 1:500), rabbit anti-PAX6 (Covance, PRB-278P, 1:200), rat anti-KRT8/TROMA-I (DSHB, AB 531826, 1:200), rabbit anti-CDX2 (Abcam, ab76541, 1:400), mouse anti-Nestin (Merk, MAB353, 1:200), rat anti-GATA4 (Invitrogen, 14998082, 1:500), rabbit anti-CDKN1C (Abcam, ab75974, 1:250), rat anti-Mki67 (Invitrogen (eBioscience), 14-5698-82, 1:200), mouse anti-ASCL1 (Invitrogen (eBioscience), 14-5794-82, 1:200), rabbit anti-TPBPA (Abcam, ab104401, 1:200). Coverslips were washed three times with PBS 5 min at room temperature while gently rocking, and then incubated with secondary antibody in blocking solution for 1 h at room temperature. DAPI (0.5 µg/µl) was added together with a secondary antibody. Secondary antibodies used: goat anti-Mouse-488 (Invitrogen, A11029, 1:1000), goat anti-Rabbit-488 (Invitrogen, A11034, 1:1000), goat anti-Rat-488 (Invitrogen, A11006, 1:1000), goat anti-Mouse-568 (Invitrogen, A11031, 1:1000), goat anti-Rabbit-568 (Invitrogen, A11036, 1:1000), goat anti-Mouse-647 (Invitrogen, A21247, 1:1000), goat anti-Rabbit-647 (Invitrogen, A21236, 1:1000), goat anti-Rat-647 (Invitrogen, A21245, 1:1000). Coverslips were then washed as before and mounted on glass using Prolong™ Glass Antifade Mountant (Invitrogen, P36984).

For flow cytometry experiments with intracellular immunostaining of cells (FACS-IF), cells were grown as described above, dissociated with trypsin, washed with PBS and fixed for 15 min at room temperature with 4% PFA in PBS. Cells were then washed with PBS and 50–100 µl of cell suspension were pelleted at 400×*g* for 5 min,

permeabilized and blocked with 100 µl blocking solution for 1 h at room temperature. Cells were then centrifuged at 400×*g* for 5 min and washed twice with 200 µl PBS. Cells were incubated with primary antibody (same as for microscopy samples) in the blocking solution for 1 h at room temperature, washed twice with 200 µl PBS with 400×*g* 5 min centrifugation in between, and incubated with secondary antibody (same as for microscopy samples) and DAPI (0.5 µg/µl) in the blocking solution for 1 h at room temperature. Cells were washed twice with 200 µl PBS and resuspended in 100 µl PBS and quantified using FACS BD LSR Fortessa (BD Biosciences).

RT-qPCR

Total RNA was extracted using the Qiagen RNeasy mini kit. Then, 1–2 µg of total RNA was reverse transcribed using the SuperScript™ III Reverse Transcriptase (Invitrogen). RT-qPCR was performed using GoTaq® qPCR MasterMix (Promega). RT-qPCR primer sequences used in this study can be found in Supplementary Table 3. Relative RNA levels were calculated from C_t values by the standard $\Delta\Delta C_t$ method and normalized to *Actin* mRNA levels.

Screen

Retroviral CRISPR-UMI sgRNA library was produced by transfection of the PlatinumE cells (Cell Biolabs) as described before[44]. In the morning, 3×10^8 ESC were plated on the gelatin-coated plates (10^7 cells per plate) without feeders and infected with a 10% infection rate 1 h later. In the evening, feeders were seeded on top of the ESC. Successfully infected cells were selected with 0.5 mg/ml G418 (Gibco, 11811-031) for 5 days changing medium daily and splitting while maintaining 500x library coverage at all times. Nine days post-infection, 5×10^6 cells per plate were plated on Matrigel-coated 15 cm plates in the morning and induced with N2B27+Dox medium in the evening with a total of 6×10^7 cells per condition. Half of the medium was exchanged daily. From day 4 onward, iN were purified by the addition of 4 µM cytosine β-D-arabinofuranoside (AraC, Sigma-Aldrich) and 1 µg/ml puromycin until day 6 post-induction. For the iNSC samples, cells were trypsinized and expanded after day 3 post-induction for additional 4 days in N2B27+Dox medium containing 20 ng/ml FGF2 (PeproTech, 450-33) and 20 ng/ml EGF (PeproTech, 315-09). iT were enriched by plating cells at day 6 in ESCM after trypsinization onto gelatin-coated 15 cm plates for 20 min and then washing off unattached cells with PBS. ESC were kept in parallel in ESCM medium and feeders as described above. At the end of differentiation, cells were trypsinized, washed with PBS, pelleted by centrifugation at 300×*g* for 5 min and frozen. Lastly, gDNA isolation, PacI digestion (NEB), size selection, PCR amplification and NGS were carried out as previously published[44,78].

RNAseq

Cells were collected at 0 h, 24 h and 48 h post-induction as bulk samples, and on days 3, 4, and 6, cells were sorted into Venus-positive and Venus-negative cells using FACSAria III (BD bioscience). Cells were washed with PBS and processed using the Qiagen RNeasy mini kit. RNA was additionally treated with a TURBO DNA-free™ kit (Invitrogen) according to the manufacturer's protocols. The integrity of RNA was measured using Fragment Analyzer™ (Advanced Analytical). RNAseq-Libraries were prepared using QuantSeq 3′ mRNA-Seq Library Prep Kit (Lexogen GmbH). LUTHOR 3′-scRNAseq (Lexogen GmbH) libraries were prepared using the manufacturer's protocol after sorting cells into 5 µl lysis agent. Concentrations and distributions of the libraries were checked with the Fragment Analyzer™ using HS NGS Fragment Kit (Agilent; DNF-474-0500). Libraries were then pooled and sequenced using Illumina Nextseq550 in a single read 75 cycles run.

ChIPseq

Two replicates of 25×10^6 E14 (CAG-rtTA-Hygro, LV-TetOL-Flag-Ascl1/Ngn2-T2A-Puro) cells were collected 24 h after induction, washed with PBS and fixed with 1% freshly prepared formaldehyde (FA) solution in PBS for 7 min at room temperature. FA was then quenched with 0.125 M glycine. Nuclei were extracted by cell lysis in 50 mM HEPES-KOH pH 8.0, 140 mM NaCl, 1 mM EDTA, 10 % glycerol, 0.5% NP40, 0.25 % Triton X-100 for 10 min at 4 °C, followed by 5 min on ice of 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl. Nuclei were then washed and resuspended in 10 mM Tris-HCl pH 8.0, 0.1 % SDS, 1 mM EDTA with 1x Complete mini protease inhibitors (Roche), and chromatin was sheared using Covaris E220 High Performance Focused Ultrasonicator (Duty factor 5.0, PIP 140.0, 200 cycles per burst at 4 °C). Fragment sizes of 250–800 bp were confirmed via agarose gel electrophoresis. 1% of fragmented chromatin was kept as an input sample. For each replicate, a duplicate of 50 µg of chromatin was used for overnight 4 °C incubation with 5 µg of anti-FLAG (Sigma, F1804) in IP Buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% DOC, 0.1% SDS) with BSA blocked Protein G coupled Dynabeads (Thermo Fisher Scientific; Blocking was done for 3 h at 4 °C). Beads were then washed eight times with IP buffer and once with TE, 50 mM NaCl. The DNA was eluted twice with 150 µl of 1% SDS, 0.1 M $NaHCO_3$ for 20 min at 65 °C. The eluate was then treated with RNase A and Proteinase K and reverse crosslinked overnight at 65 °C. DNA was purified using phenol/chloroform extraction and isopropanol precipitation. ChIP libraries were prepared using the NEBNext Ultra-II kit (New England Biolabs) following the manufacturer's protocol. The quality of the libraries was assessed with the Fragment Analyzer™ using HS NGS Fragment Kit, and libraries were sequenced using Illumina HiSeq V4 in a single read 50 bp sequencing run.

Data analysis

RNAseq. RNA-seq reads were trimmed using BBDuk v38.06 (ref=polyA.fa.gz,truseq.fa.gz k = 13 ktrim=r useshortkmers=t mink=5 qtrim=r trimq=10 minlength=20) and reads mapping to abundant sequences included in the iGenomes Ensembl GRCm38 bundle (mouse rDNA, mouse mitochondrial chromosome, phiX174 genome, adapter) were removed using bowtie2 v2.3.4.1 alignment. The remaining reads were analyzed using genome and gene annotation for the GRCm38/mm10 assembly obtained from *Mus musculus* Ensembl release 94. Reads were aligned to the genome using star v2.6.0c, and reads in genes were counted with featureCounts (subread v1.6.2) using strand-specific read counting for QuantSeq experiments (-s 1). For differential expression analysis and variance stabilized transformation for PCA analysis, we used R package Deseq2[79] (v.1.32.0). Significantly differentially expressed genes (DEG) were considered those genes with FDR < 0.05 and −1 <Log2(Fold change) <1. For the KEGG pathway enrichment analysis and identification of the alternative cell lineages using the PanglaoDB database[32], we used Enrichr[80]. For the DEG network analysis and visualization, we used the STRING database and Cytoscape (v.3.8.0).

ChIPseq. ChIP-seq reads were trimmed using trim-galore v0.4.4 and thereafter aligned to the mm10 reference genome using bwa mem v0.7.17. Duplicated reads were removed using Picard MarkDuplicates (v2.23.4.). Complexity and overall data quality were assessed using phantompeakqualtools and deepTools plotFingerprint (v 3.5.0). To generate bigwig files from corresponding bam files, we used deepTools bamCoverage (v 3.5.0), with parameter -normalizeUsing RPKM. Peak calling from the sorted BAM files was done using MACS2 (v2.1.1). Peaks in blacklist regions were identified using bedtools intersect (v2.27.1) and mm10.blacklist.bed.gz v1. To analyze overlapping and unique peaks between Ascl1 and Ngn2, as well as plotting peak heatmaps, we used the R package Diffbind[81] (v3.2.7) with minOverlap = 4 to collapse replicates after making a consensus peaksets. For determining binding motifs, we used MEME-ChIP (v5.1.1) with -meme-minw 6 -meme-maxw 15 parameters. For

plotting the binding profiles, we combined the reads from the replicates using UCSC bigWigMerge and used R package karyoploteR (v1.18.0).

CRISPR-Cas9 screen analysis. Analysis of CRISPR-Cas9 screens was done according to previously published protocol[44,78]. In short, reads were trimmed to 20 nt sgRNA sequence using fastx-toolkit (v0.0.14) and mapped to reference sequences using bowtie (v1.1.2). Experimental indices and mapped reads were collapsed into count tables using in-house Python scripts. Gene enrichment was determined using MAGeCK[82] (v0.5.4). Visualization was done using the R package ggplot2 (v3.3.6).

Statistics and reproducibility. Experiments were independently performed at least three times (unless otherwise stated). Figure 3g screen validation experiment, a replicate was excluded for Setdlb due to an inefficient number of cells, thus resulting in poor sample quantification. No key conclusions were drawn based on this sample. Results are reported as mean ± SD. For comparing between two groups, a two-tailed unpaired Student's t-test was performed. One-way ANOVA followed by Dunnett's multiple comparison test was applied for comparisons of multiple groups. Statistical tests were performed using R. Only representative micrographs are shown from at least two independent replicates, as similar results were obtained between replicates. In Supplementary Figs. 4, 12c–d, micrographs of all the replicates are shown.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw NGS data produced in this study have been deposited in Gene Expression Omnibus (GEO) database under super series accession number GSE206872: ChIP-seq (GSE206869), RNA-seq (GSE206870, GSE208199), CRISPR-Cas9 screen (GSE206871). Single-cell RNA-seq data for reanalysis can be obtained from GSE125620[6]. Source data are provided with this paper.

Code availability

Only standard bioinformatic code was used for the data analysis and is available on request.

References

1. Balsalobre, A. & Drouin, J. Pioneer factors as master regulators of the epigenome and cell fate. *Nat. Rev. Mol. Cell Biol.* **23**, 449–464 (2022).
2. Bergsland, M. et al. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* **25**, 2453–2464 (2011).
3. Buecker, C. et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
4. Holmberg, J. & Perlmann, T. Maintaining differentiated cellular identity. *Nat. Rev. Genet.* **13**, 429–439 (2012).
5. Li, M. & Belmonte, J. C. I. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* **18**, 180–191 (2017).
6. Aydin, B. et al. Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nat. Neurosci.* **22**, 897–908 (2019).
7. Wapinski, O. L. et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621–635 (2013).
8. Ali, F. R. et al. The phosphorylation status of Ascl1 is a key determinant of neuronal differentiation and maturation in vivo and in vitro. *Development* **141**, 2216–2224 (2014).
9. Azzarelli, R. et al. ASCL1 phosphorylation and ID2 upregulation are roadblocks to glioblastoma stem cell differentiation. *Sci. Rep.* **12**, 2341 (2022).
10. Blomfield, I. M. et al. Id4 promotes the elimination of the pro-activation factor Ascl1 to maintain quiescence of adult hippocampal stem cells. *eLife* **8**, e48561 (2019).
11. Han, S. et al. Proneural genes define ground-state rules to regulate neurogenic patterning and cortical folding. *Neuron* **109**, 2847–2863.e11 (2021).
12. Le Dréau, G. et al. E proteins sharpen neurogenesis by modulating proneural bHLH transcription factors' activity in an E-box-dependent manner. *eLife* **7**, e37267 (2018).
13. Viñals, F. et al. BMP-2 decreases Mash1 stability by increasing Id1 expression. *EMBO J.* **23**, 3527–3537 (2004).
14. Herrero-Navarro, Á. et al. Astrocytes and neurons share region-specific transcriptional signatures that confer regional identity to neuronal reprogramming. *Sci. Adv.* **7**, eabe8978 (2021).
15. Urbán, N. et al. Return to quiescence of mouse neural stem cells by degradation of a proactivation protein. *Science* **353**, 292–295 (2016).
16. Lee, Q. Y. et al. Pro-neuronal activity of Myod1 due to promiscuous binding to neuronal genes. *Nat. Cell Biol.* **22**, 401–411 (2020).
17. Nishiyama, A. et al. Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. *Cell Stem Cell* **5**, 420–433 (2009).
18. Schiebinger, G. et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943.e22 (2019).
19. Treutlein, B. et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**, 391–395 (2016).
20. Briggs, J. A. et al. Mouse embryonic stem cells can differentiate via multiple paths to the same state. *eLife* **6**, e26945 (2017).
21. Lin, H.-C. et al. NGN2 induces diverse neuron types from human pluripotency. *Stem Cell Rep.* **16**, 2118–2127 (2021).
22. Mertens, J., Marchetto, M. C., Bardy, C. & Gage, F. H. Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nat. Rev. Neurosci.* **17**, 424–437 (2016).
23. Wang, H., Yang, Y., Liu, J. & Qian, L. Direct cell reprogramming: approaches, mechanisms and progress. *Nat. Rev. Mol. Cell Biol.* **22**, 410–424 (2021).
24. Bocchi, R., Masserdotti, G. & Götz, M. Direct neuronal reprogramming: fast forward from new concepts toward therapeutic approaches. *Neuron* **110**, 366–393 (2022).
25. Liu, Y. et al. CRISPR activation screens systematically identify factors that drive neuronal fate and reprogramming. *Cell Stem Cell* **23**, 758–771.e8 (2018).
26. Masserdotti, G. et al. Transcriptional mechanisms of proneural factors and REST in regulating neuronal reprogramming of astrocytes. *Cell Stem Cell* **17**, 74–88 (2015).
27. Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
28. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
29. Parras, C. M. et al. Divergent functions of the proneural genes *Mash1* and *Ngn2* in the specification of neuronal subtype identity. *Genes Dev.* **16**, 324–338 (2002).
30. Yang, N. et al. Generation of pure GABAergic neurons by transcription factor programming. *Nat. Methods* **14**, 621–628 (2017).
31. Velasco, S. et al. A multi-step transcriptional and chromatin state cascade underlies motor neuron programming from embryonic stem cells. *Cell Stem Cell* **20**, 205–217.e8 (2017).
32. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**, baz046 (2019).

33. Knöfler, M. et al. Human placenta and trophoblast development: key molecular mechanisms and model systems. *Cell. Mol. Life Sci.* **76**, 3479–3496 (2019).
34. Casey, B. H., Kollipara, R. K., Pozo, K. & Johnson, J. E. Intrinsic DNA binding properties demonstrated for lineage-specifying basic helix-loop-helix transcription factors. *Genome Res.* **28**, 484–496 (2018).
35. Guillemot, F., Nagy, A., Auerbach, A., Rossant, J. & Joyner, A. L. Essential role of Mash-2 in extraembryonic development. *Nature* **371**, 333–336 (1994).
36. Hughes, M. et al. The Hand1, Stra13 and Gcm1 transcription factors override FGF signaling to promote terminal differentiation of trophoblast stem cells. *Dev. Biol.* **271**, 26–37 (2004).
37. Kinoshita, M. et al. Disabling de novo DNA methylation in embryonic stem cells allows an illegitimate fate trajectory. *Proc. Natl Acad. Sci. USA* **118**, e2109475118 (2021).
38. Nehme, R. et al. Combining NGN2 programming with developmental patterning generates human excitatory neurons with NMDAR-mediated synaptic transmission. *Cell Rep.* **23**, 2509–2523 (2018).
39. Hitoshi, S. et al. Notch pathway molecules are essential for the maintenance, but not the generation, of mammalian neural stem cells. *Genes Dev.* **16**, 846–858 (2002).
40. Pierfelice, T., Alberi, L. & Gaiano, N. Notch in the vertebrate nervous system: an old dog with new tricks. *Neuron* **69**, 840–855 (2011).
41. Tian, R. et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived. *Neurons Neuron* **104**, 239–255.e12 (2019).
42. Tracey, T. J., Steyn, F. J., Wolvetang, E. J. & Ngo, S. T. Neuronal lipid metabolism: multiple pathways driving functional outcomes in health and disease. *Front. Mol. Neurosci.* **11**, 10 (2018).
43. Hulme, A. J., Maksour, S., St-Clair Glover, M., Miellet, S. & Dottori, M. Making neurons, made easy: the use of Neurogenin-2 in neuronal differentiation. *Stem Cell Rep.* **17**, 14–34 (2022).
44. Michlits, G. et al. CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nat. Methods* **14**, 1191–1197 (2017).
45. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
46. Gascón, S. et al. Identification and successful negotiation of a metabolic checkpoint in direct neuronal reprogramming. *Cell Stem Cell* **18**, 396–409 (2016).
47. Russo, G. L. et al. CRISPR-mediated induction of neuron-enriched mitochondrial proteins boosts direct glia-to-neuron conversion. *Cell Stem Cell* **28**, 524–534.e7 (2021).
48. Cole, M. F., Johnstone, S. E., Newman, J. J., Kagey, M. H. & Young, R. A. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev.* **22**, 746–755 (2008).
49. Jin, L. et al. Serine threonine kinase receptor-associated protein deficiency impairs mouse embryonic stem cells lineage commitment through CYP26A1-mediated retinoic acid homeostasis. *Stem Cells* **36**, 1368–1379 (2018).
50. Kielman, M. F. et al. Apc modulates embryonic stem-cell differentiation by controlling the dosage of β-catenin signaling. *Nat. Genet.* **32**, 594–605 (2002).
51. Quan, Y. et al. Cnot8 eliminates naïve regulation networks and is essential for naïve-to-formative pluripotency transition. *Nucleic Acids Res.* **50**, 4414–4435 (2022).
52. Wray, J. et al. Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat. Cell Biol.* **13**, 838–845 (2011).
53. Ying, Q.-L. et al. The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
54. Zhang, Y., Ding, H., Wang, X. & Ye, S.-D. Modulation of STAT3 phosphorylation by PTPN2 inhibits naïve pluripotency of embryonic stem cells. *FEBS Lett.* **592**, 2227–2237 (2018).
55. Dai, Q. et al. Cyclin K-containing kinase complexes maintain self-renewal in murine embryonic stem cells. *J. Biol. Chem.* **287**, 25344–25352 (2012).
56. Lei, T. et al. Cyclin K regulates prereplicative complex assembly to promote mammalian cell proliferation. *Nat. Commun.* **9**, 1876 (2018).
57. Dutto, I., Tillhon, M., Cazzalini, O., Stivala, L. A. & Prosperi, E. Biology of the cell cycle inhibitor p21CDKN1A: molecular mechanisms and relevance in chemical toxicology. *Arch. Toxicol.* **89**, 155–178 (2015).
58. Johnson, P. F. Molecular stop signs: regulation of cell-cycle arrest by C/EBP transcription factors. *J. Cell Sci.* **118**, 2545–2555 (2005).
59. Bates, L. E., Alves, M. R. P. & Silva, J. C. R. Auxin-degron system identifies immediate mechanisms of OCT4. *Stem Cell Rep.* **16**, 1818–1831 (2021).
60. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372–376 (2000).
61. Hough, S. R., Clements, I., Welch, P. J. & Wiederholt, K. A. Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. *Stem Cells* **24**, 1467–1475 (2006).
62. Mall, M. et al. Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature* **544**, 245–249 (2017).
63. Pereira, L., Yi, F. & Merrill, B. J. Repression of Nanog gene transcription by Tcf3 limits embryonic stem cell self-renewal. *Mol. Cell. Biol.* **26**, 7479–7491 (2006).
64. Yi, F. et al. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. *Nat. Cell Biol.* **13**, 762–770 (2011).
65. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).
66. Creff, J. & Besson, A. Functional versatility of the CDK inhibitor p57Kip2. *Front. Cell Dev. Biol.* **8**, 584590 (2020).
67. Li, N. et al. p57 suppresses the pluripotency and proliferation of mouse embryonic stem cells by positively regulating p53 activation. *Stem Cells Int.* **2021**, 1–13 (2021).
68. Cheloufi, S. et al. The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).
69. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
70. Lu, Y.-L. & Yoo, A. S. Mechanistic insights into MicroRNA-induced neuronal reprogramming of human adult fibroblasts. *Front. Neurosci.* **12**, 522 (2018).
71. Wapinski, O. L. et al. Rapid chromatin switch in the direct reprogramming of fibroblasts to neurons. *Cell Rep.* **20**, 3236–3247 (2017).
72. Marro, S. et al. Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* **9**, 374–382 (2011).
73. Colasante, G., Rubio, A., Massimino, L. & Broccoli, V. Direct neuronal reprogramming reveals unknown functions for known transcription factors. *Front. Neurosci.* **13**, 283 (2019).
74. Smith, D. K., Yang, J., Liu, M.-L. & Zhang, C.-L. Small molecules modulate chromatin accessibility to promote NEUROG2-mediated fibroblast-to-neuron reprogramming. *Stem Cell Rep.* **7**, 955–969 (2016).
75. Liu, M.-L. et al. Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. *Nat. Commun.* **4**, 2183 (2013).
76. Chanda, S. et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Rep.* **3**, 282–296 (2014).
77. Tsunemoto, R. et al. Diverse reprogramming codes for neuronal identity. *Nature* **557**, 375–380 (2018).

78. Michlits, G., Burkard, T. R., Novatchkova, M. & Elling, U. CRISPR-UMI step by step: a protocol for robust CRISPR screening. *Protoc. Exch.* https://doi.org/10.1038/protex.2017.111 (2017).
79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
80. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
81. Stark, R. & Brown, G. DiffBind: differential binding analysis of ChIP-seq peak data. Bioconductor. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf (2011).
82. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

Acknowledgements

We thank the VBCF-NGS team (https://www.vbcf.ac.at) for providing deep sequencing services and the IMBA-IMP BioOptics team for sharing expertise in flow cytometry, microscopy and cell sorting. We thank all the members of the Elling lab for their input and scientific discussions; Noelia Urban, Nicolas Rivron, and Bonkyoung Koo labs for sharing reagents; Dr. Torsten Reda and Dr. Mantas Survila for their expertise and LUTHOR 3′-scRNAseq library preparations.

Author contributions

U.E., M.W. and G.V. conceived and planned this project. G.V. designed and conducted experiments with the help of C.R., J.C.B. and J.L. G.V. performed data analysis with the help of M.N. and G.M. R.Y. provided expertise and helped with ChIP-seq experiments. M.W. provided initial cell lines for the experiments and valuable guidance throughout the project. U.E. supervised the project. G.V. and U.E. wrote the manuscript with input from all the co-authors.

Competing interests

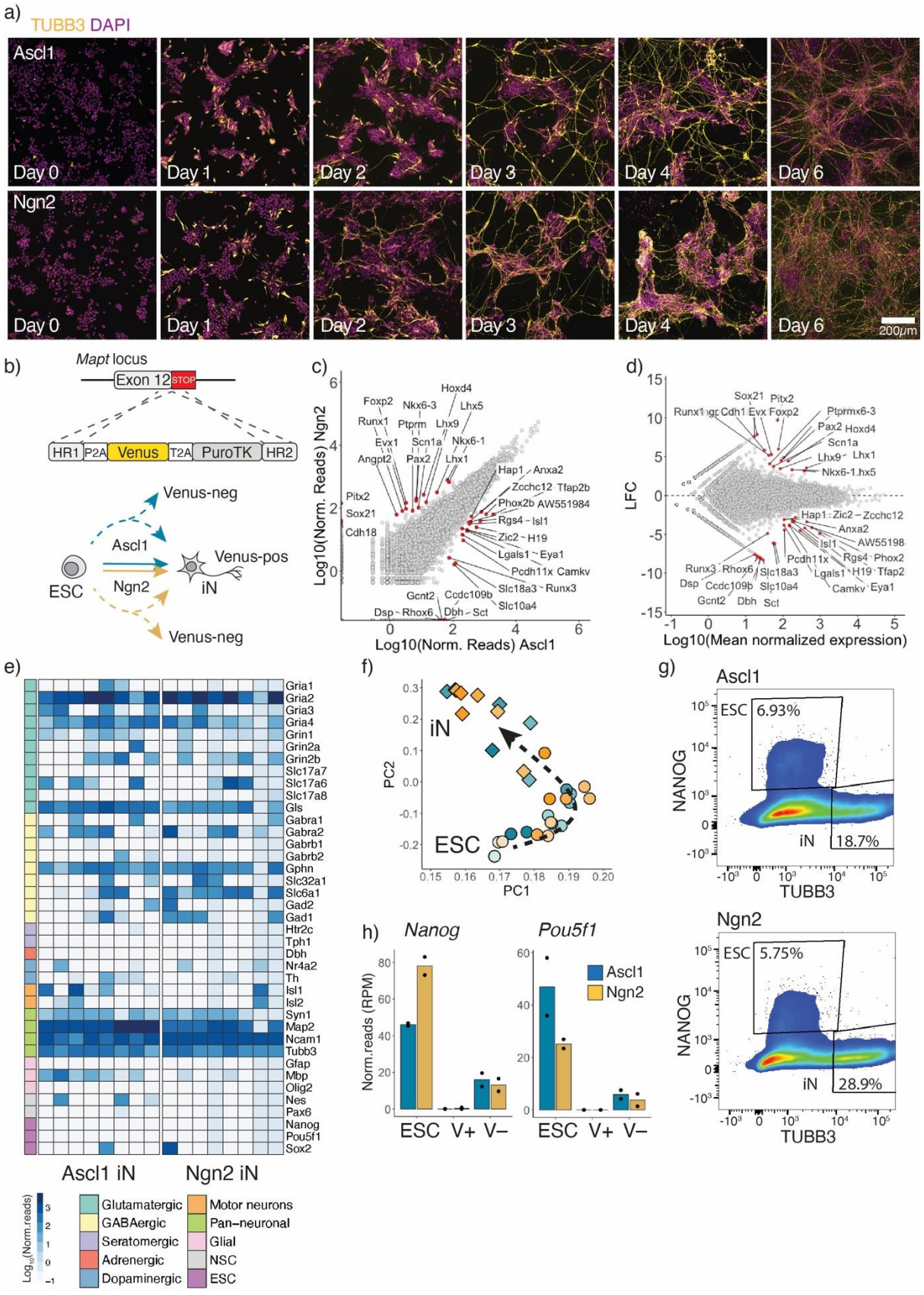
The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-40803-y.

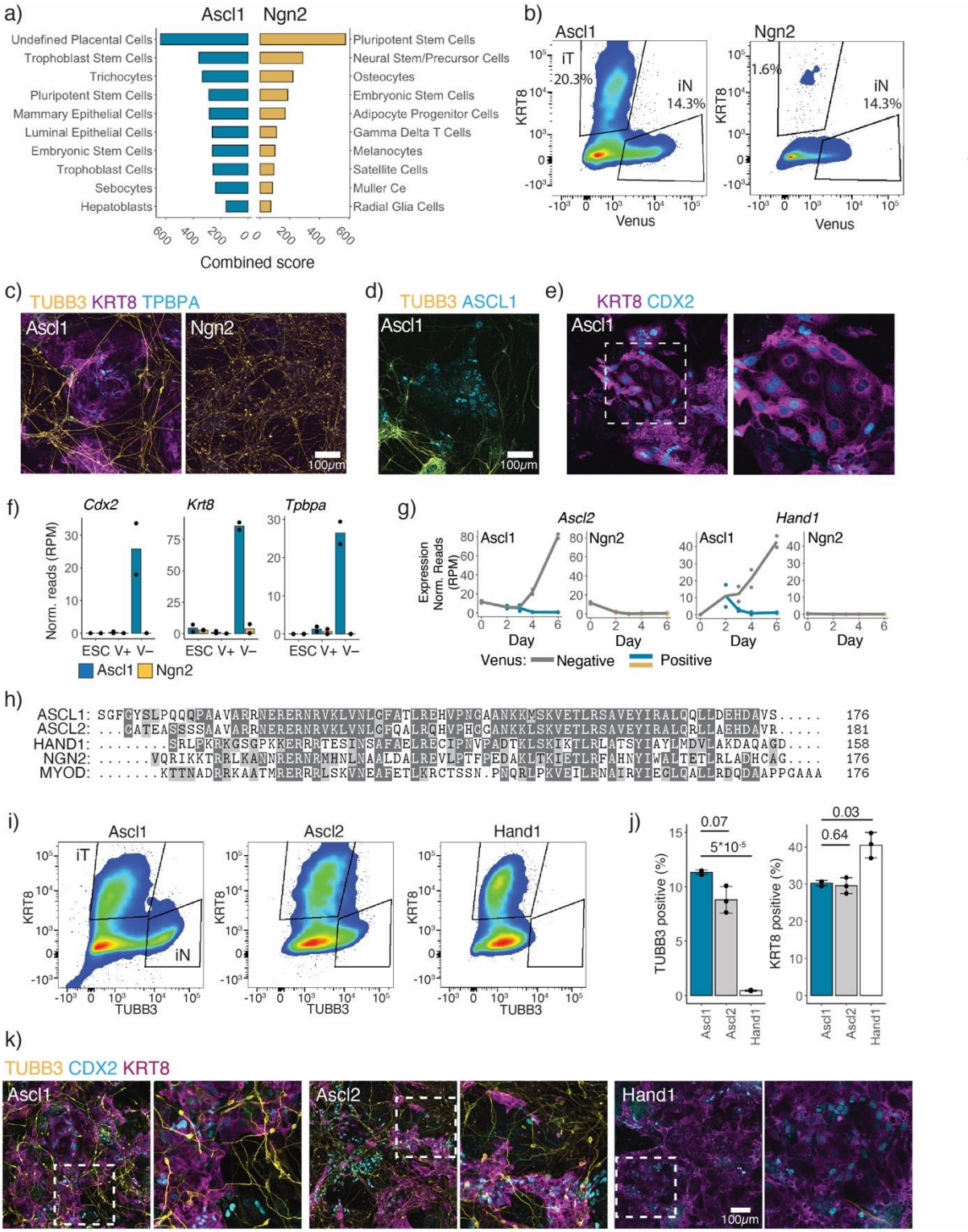
Correspondence and requests for materials should be addressed to Gintautas Vainorius or Ulrich Elling.

Peer review information *Nature Communications* thanks Raul Bressan and the other anonymous reviewer(s) for their contribution to the peer review of this work.


a)
TUBB3 DAPI
Ascl1
Day 0
Day 1
Day 2
Day 3
Day 4
Day 6
Ngn2
Day 0
Day 1
Day 2
Day 3
Day 4
Day 6
200μm
b)
Mapt locus
Exon 12
STOP
HR1
P2A
Venus
T2A
PuroTK
HR2
Venus-neg
Ascl1
Venus-pos
ESC
Ngn2
iN
Venus-neg
c)
Log10(Norm. Reads) Ngn2
Log10(Norm. Reads) Ascl1
Hoxd4
Foxp2
Nkx6-3
Lhx5
Runx1
Ptprm
Lhx9
Evx1
Scn1a
Nkx6-1
Angpt2
Pax2
Lhx1
Hap1
Anxa2
Zcchc12
Tfap2b
Pitx2
Phox2b
AW551984
Sox21
Rgs4
Isl1
Cdh18
Zic2
H19
Lgals1
Eya1
Pcdh11x
Camkv
Slc18a3
Runx3
Gcnt2
Ccdc109b
Slc10a4
Dsp
Rhox6
Dbh
Sct
d)
LFC
Log10(Mean normalized expression)
Sox21
Pitx2
Runx1
Cdh1
Evx
Foxp2
Ptprmx6-3
Pax2
Hoxd4
Scn1a
Lhx9
Lhx1
Nkx6-1.hx5
Hap1
Zic2
Zcchc12
Anxa2
Isl1
AW55198
Pcdh11x
Rgs4
Phox2
Runx3
Rhox6
Slc18a3
Lgals1
H19
Tfap2
Dsp
Ccdc109b
Slc10a4
Camkv
Eya1
Gcnt2
Dbh
Sct
e)
Gria1
Gria2
Gria3
Gria4
Grin1
Grin2a
Grin2b
Slc17a7
Slc17a6
Slc17a8
Gls
Gabra1
Gabra2
Gabrb1
Gabrb2
Gphn
Slc32a1
Slc6a1
Gad2
Gad1
Htr2c
Tph1
Dbh
Nr4a2
Th
Isl1
Isl2
Syn1
Map2
Ncam1
Tubb3
Gfap
Mbp
Olig2
Nes
Pax6
Nanog
Pou5f1
Sox2
Ascl1 iN
Ngn2 iN
Log10(Norm.reads)
Glutamatergic
GABAergic
Seratomergic
Adrenergic
Dopaminergic
Motor neurons
Pan–neuronal
Glial
NSC
ESC
f)
PC2
PC1
iN
ESC
g)
Ascl1
ESC
6.93%
iN
18.7%
NANOG
TUBB3
Ngn2
ESC
5.75%
iN
28.9%
NANOG
TUBB3
h)
Nanog
Pou5f1
Norm.reads (RPM)
Ascl1
Ngn2
ESC
V+
V–
ESC
V+
V–

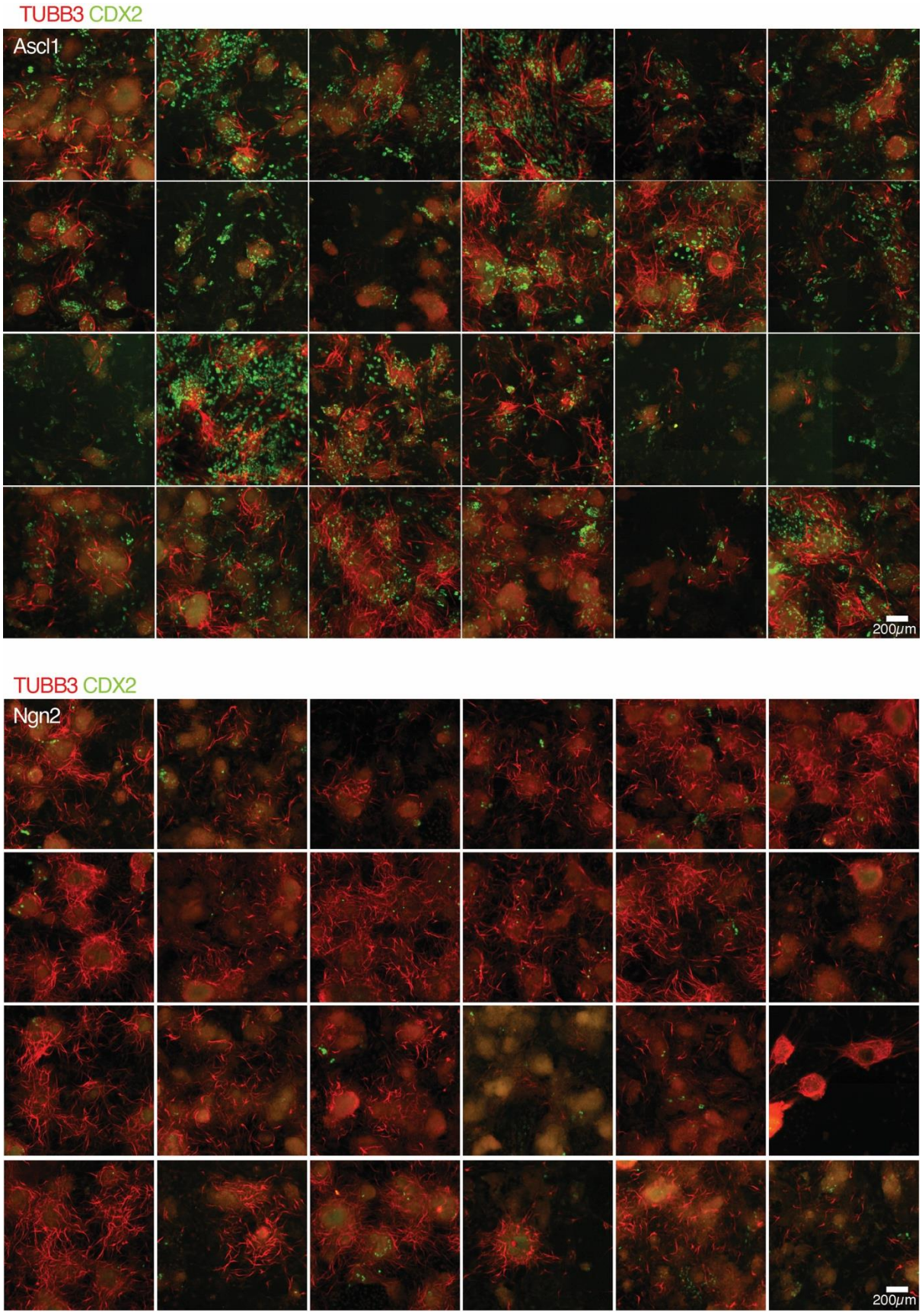
Supplementary Figure 1. Characterization of induced neurons at day 6

a, Immunostainings showing forward differentiation of ESC over time after overexpression of Ascl1 or Ngn2. **b**, Scheme of C-terminally tag of Mapt-Venus and a visual representation of ESC differentiating towards iN after Ascl1 or Ngn2 induction. Cells may take a common path (solid line) or different paths (dashed line) between similar cell states, generating to Mapt-Venus positive iN population, while failure to reprogram or generating alternative cell states will result in Mapt-Venus negative population. **c,d**, Scatter plot (**c**) and MA plot (**d**) comparing gene expression at Day 6 between Ascl1 and Ngn2 Venus positive cells (Fig. 1b). Differentially expressed genes are indicated in red (FDR < 0.05). **e**, Heatmap showing neuronal subtype specific markes expressed in individual neurons using LUTHOR 3' ultra-high sensitivity scRNAseq. **f**, Combined principal component analysis of time resolved bulk RNAseq (Fig. 1 c,d). **g**, FACS-IF plots showing intracellular immunostaining of the cells for pan-neuronal marker TUBB3 and core pluripotency marker NANOG at day 6. **h**, Expression of pluripotency genes at day 6 in the mRNA dataset in the Fig. 1a. Bar plot shows mean of n=2 independent biological replicates. V – Mapt-Venus expression.

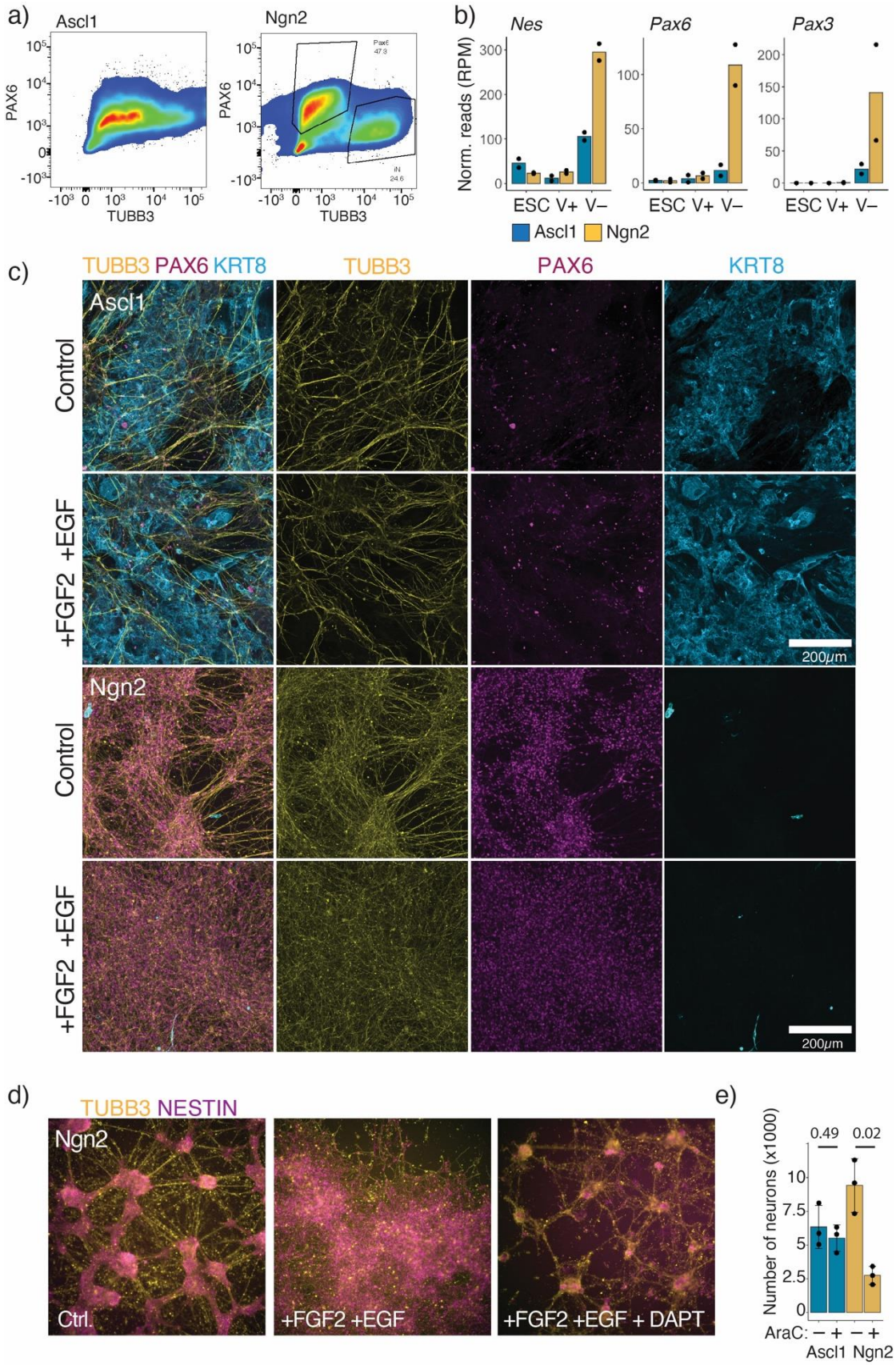

a)
Ascl1
Ngn2
Undefined Placental Cells
Trophoblast Stem Cells
Trichocytes
Pluripotent Stem Cells
Mammary Epithelial Cells
Luminal Epithelial Cells
Embryonic Stem Cells
Trophoblast Cells
Sebocytes
Hepatoblasts
Pluripotent Stem Cells
Neural Stem/Precursor Cells
Osteocytes
Embryonic Stem Cells
Adipocyte Progenitor Cells
Gamma Delta T Cells
Melanocytes
Satellite Cells
Muller Ce
Radial Glia Cells
600 400 200 0 0 200 400 600
Combined score
b)
Ascl1
iT
20.3%
iN
14.3%
KRT8
Venus
Ngn2
1.6%
iN
14.3%
KRT8
Venus
c)
TUBB3 KRT8 TPBPA
Ascl1
Ngn2
100μm
d)
TUBB3 ASCL1
Ascl1
100μm
e)
KRT8 CDX2
Ascl1
f)
Cdx2
Krt8
Tpbpa
Norm. reads (RPM)
ESC V+ V–
Ascl1 Ngn2
g)
Ascl2
Hand1
Ascl1
Ngn2
Expression Norm. Reads (RPM)
Day
Venus: Negative Positive
h)
ASCL1: SGFGYSLPQQQPAAVARRNERERNRVKLVNLGFATLREHVPNGAANKKMSKVETLRSAVEYIRALQQLLDEHDAVS..... 176
ASCL2: ...GATEASSSSAAVARRNERERNRVKLVNLGFQALRQHVPHGGANKKLSKVETLRSAVEYIRALQRLLAEHDAVR..... 181
HAND1: .........SRLPKRKGSGPKKERRRTESINSAFAELRECIPNVPADTKLSKIKTLRLATSYIAYLMDVLAKDAQAGD.... 158
NGN2: .......VQRIKKTRRLKANNRERNRMHNLNAALDALREVLPTFPEDAKLTKIETLRFAHNYIWALTETLRLADHCAG.... 176
MYOD: .......KTTNADRRKAATMRERRRLSKVNEAFETLKRCTSSN.PNQRLPKVEILRNAIRYIEGLQALLRDQDAAPPGAAA 176
i)
Ascl1
Ascl2
Hand1
iT
iN
KRT8
TUBB3
j)
0.07
5*10-5
0.03
0.64
TUBB3 positive (%)
KRT8 positive (%)
Ascl1
Ascl2
Hand1
k)
TUBB3 CDX2 KRT8
Ascl1
Ascl2
Hand1
100μm

Supplementary Figure 2. Identification of side population formed after Ascl1 overexpression

a, Cell type prediction using PanglaoDB from differentially upregulated genes at day 6 (Fig. 1e, f). Combined score is defined in Kuleshov et al, 2016. **b**, Cells immunostained for Mapt-Venus and KRT8, indicating neuronal and trophoblast populations, respectively, at day 6 post Ascl1 or Ngn2 induction. **c**, Cells immunostained for the trophoblast markers KRT8 and TPBPA and neuronal marker TUBB3 at day 6 post Ascl1 induction. Trophoblast markers were expressed after Ascl1 induction, and not Ngn2. **d**, Cells immunostained for the Ascl1 expression at day 6 post induction. **e**, Binucleated trophoblast like cells stained for immunostained for the trophoblast markers KRT8 and CDX2 at day 6 post induction of Ascl1. **f**, Expression of trophoblast marker genes at day 6 in the mRNA dataset in the Fig. 1a. Bar plot shows mean of n=2 biologically independent samples. V – Mapt-Venus expression. **g**, Time course (Fig. 1a) of the expression of bHLH transcription factors *Ascl2* and *Hand1* after induction of Ascl1 or Ngn2. Lines are drawn through the mean of n=2 biologically independent replicates. **h**, Comparison of the bHLH DNA binding domain amino acid sequences of the neural factors ASCL1, NGN2; trophoblast - ASCL2, HAND1; skeletal muscle – MYOD. The shades of letter highlights represent the similarities of the amino acids. **i**, Immunostaining for trophoblast marker KRT8 and neuronal marker TUBB3 at day 6 post induction of bHLH transcription factors Ascl1, Ascl2, Hand1. **j**, Quantification of iN and iT formation by cells immunostaining (**i**) for TUBB3 and KRT8, respectively, after 6 days of Ascl1, Ascl2 and Ngn2 induction. N = 3 biologically independent samples. Barplot indicates mean with ± SD. P values of the two-sided Welch two sample t test indicated above. **k**, Representative immunostained cells for trophoblast markers CDX2 and KRT8 and neuronal marker Tubb3 at day 6 post induction of pro-neural Ascl1, trophoblast lineage driving Ascl2 or Hand1 bHLH transcription factors. Neurons were induced after Ascl1 and Ascl2 expression, but not Hand1, while trophoblast cells were present upon induction of all three factors.

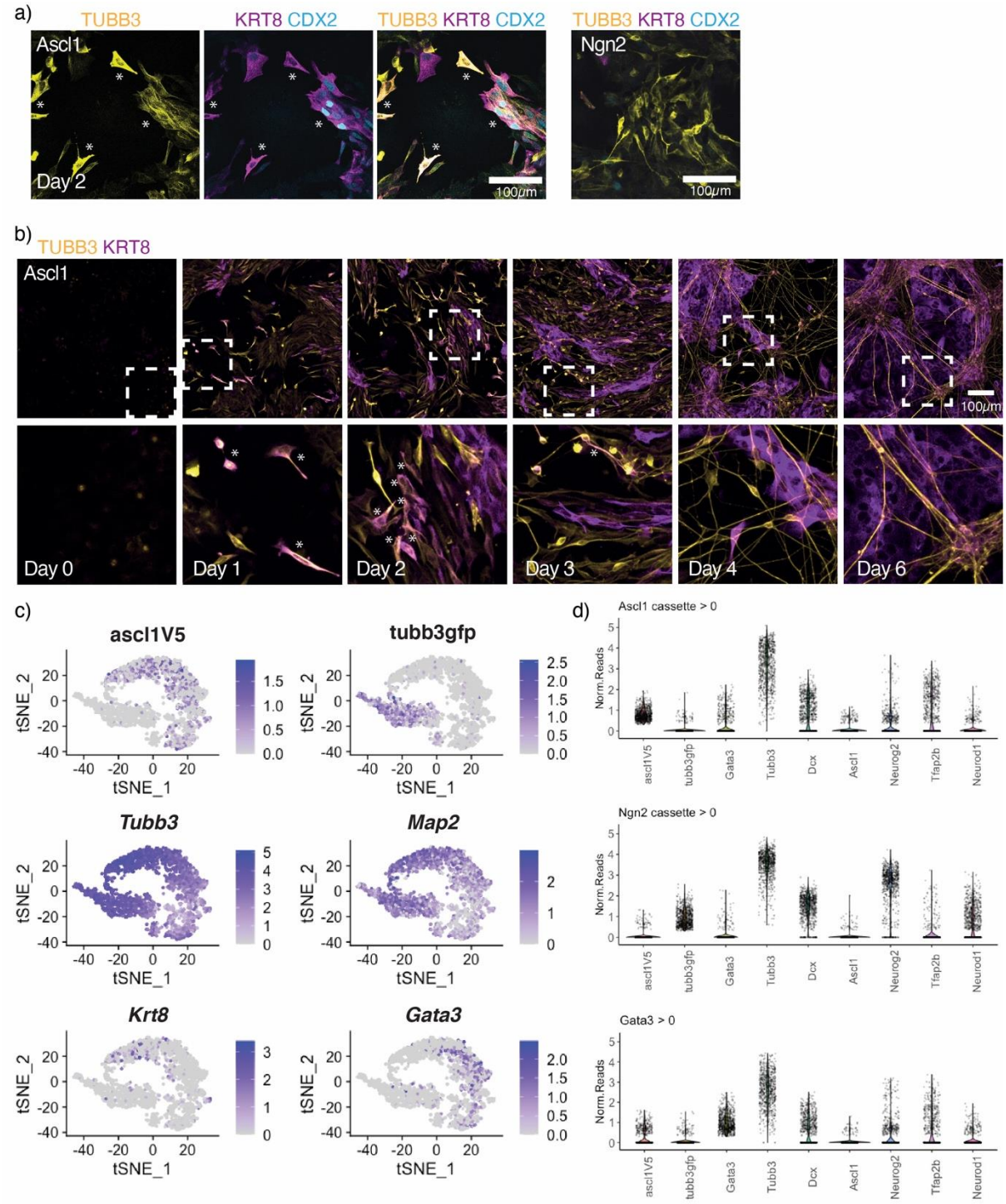


Supplementary Figure 3. Trophoblast formation in single cell derived Ascl1 or Ngn2 expressing clones
Expression of trophoblast CDX2 and neuronal TUBB3 markers in independently derived ESC clones with Ascl1 or Ngn2 overexpression cassette.


a)
Ascl1
Ngn2
PAX6
TUBB3
b)
Nes
Pax6
Pax3
Norm. reads (RPM)
ESC V+ V–
Ascl1 Ngn2
c)
TUBB3 PAX6 KRT8
TUBB3
PAX6
KRT8
Ascl1
Ngn2
Control
+FGF2 +EGF
200μm
d)
TUBB3 NESTIN
Ngn2
Ctrl.
+FGF2 +EGF
+FGF2 +EGF + DAPT
e)
0.49
0.02
Number of neurons (x1000)
AraC: – + – +
Ascl1 Ngn2

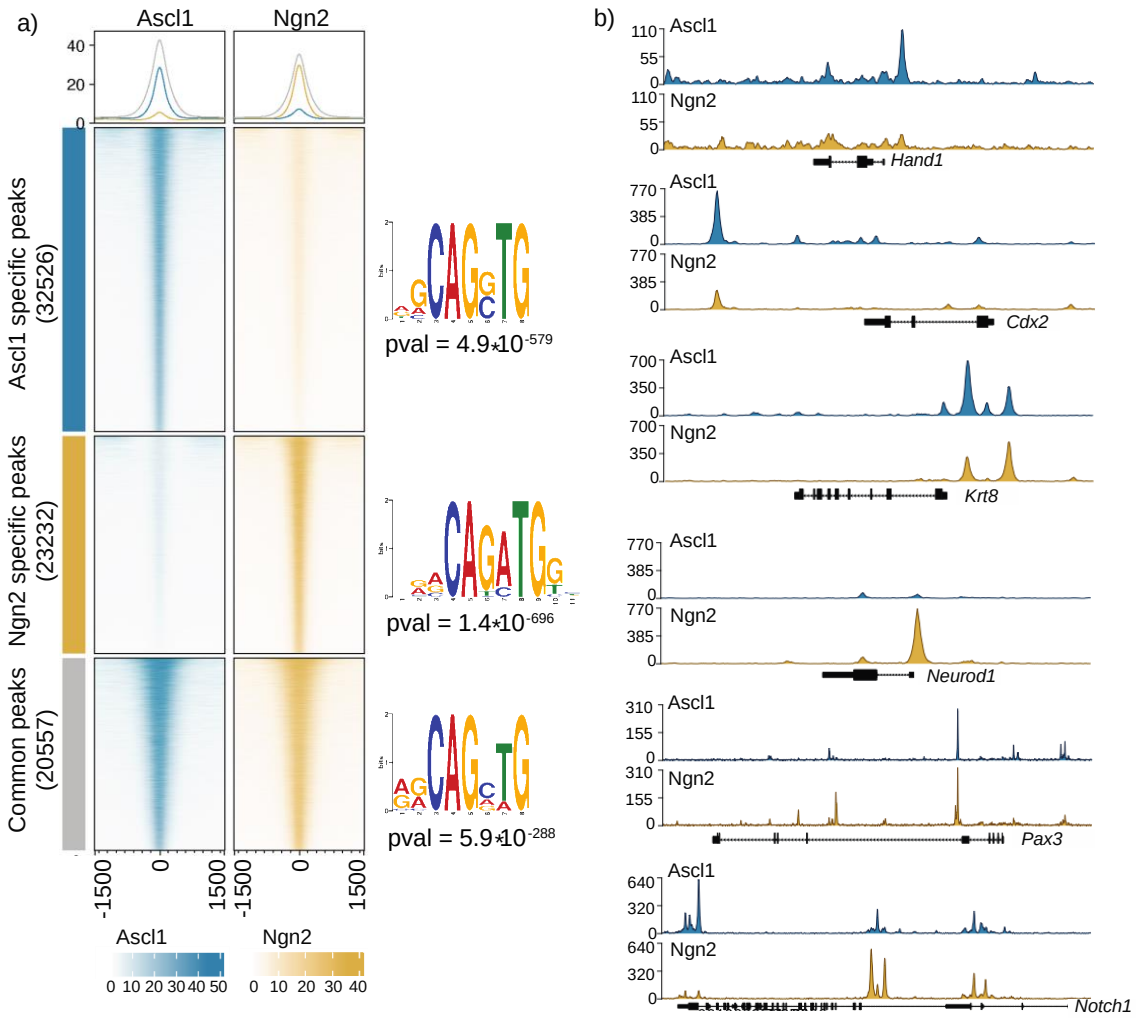
Supplementary Figure 4. Characterization of NSC-like population Ngn2 overexpression

a, Cells immunostained for TUBB3 and PAX6, indicating neuronal and NSC-like populations, respectively, after day 6 of Ascl1 or Ngn2 induction. **b**, Expression of NSC marker genes at day 6 in the mRNA dataset in the Fig. 1a. Bar plot shows mean of n=2 biologically independent samples . V – Mapt-Venus expression. **c**, Cells immunostained for neuronal marker TUBB3, trophoblast marker KRT8, NSC marker PAX6 at 6 day post induction of Ascl1 or Ngn2 in the presence of NSC inducing factors FGF2, EGF. **d**, Ngn2 induced cells immunostaining for neuronal marker TUBB3 and NSC marker NESTIN at day 6 post induction in the presence of NSC inducing factors FGF2, EGF, and Notch inhibitor DAPT. **e**, Number of neurons as measured by Mapt-Venus reporter in the presence or absence of AraC from day 4 post induction. N = 3 biologically independent samples. Bar plot shows mean with ± SD. P values of the two-sided Welch two sample t test indicated above.



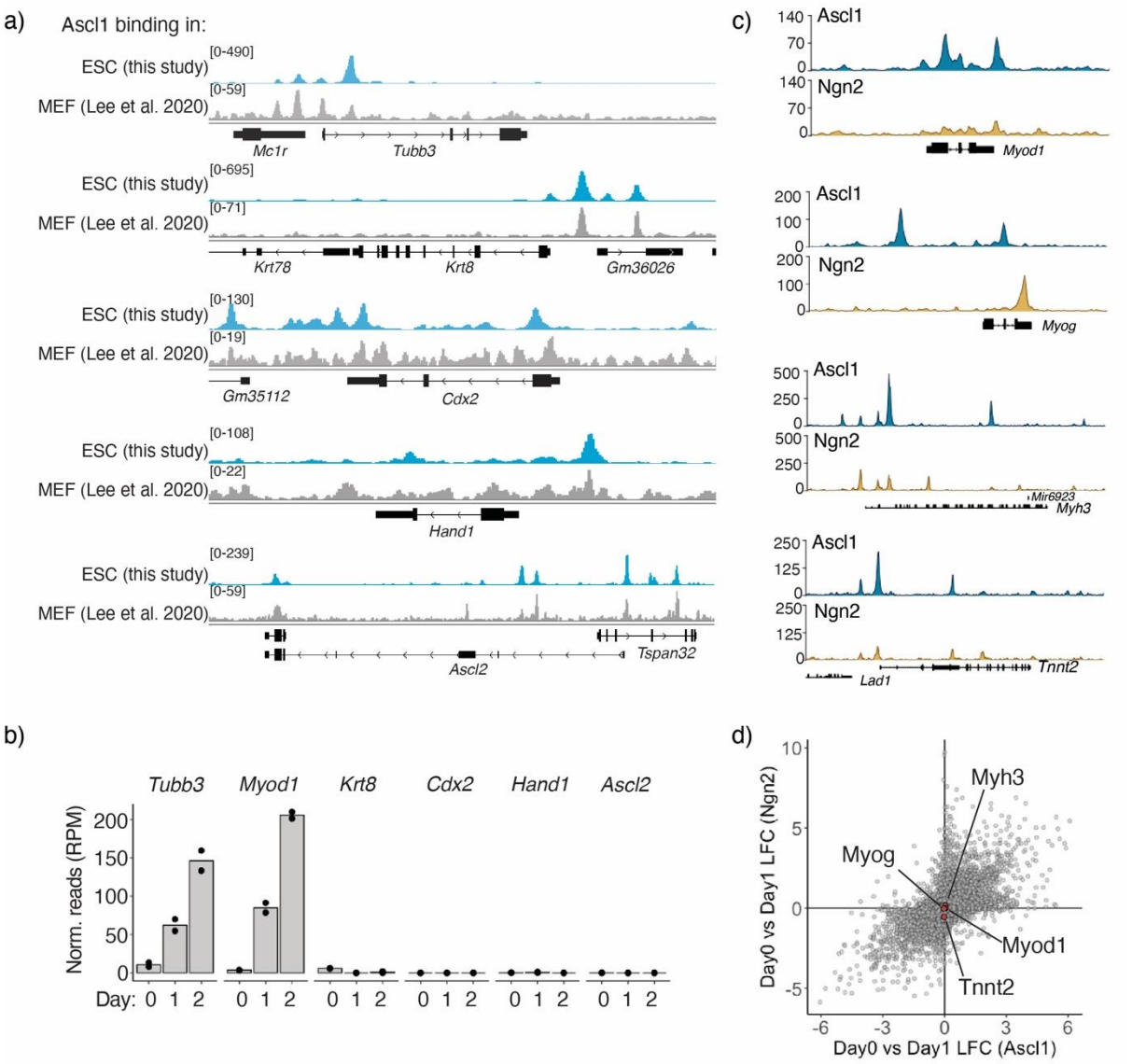
Supplementary Figure 5. iN and iT lineages coexist in early Ascl1 induced ESC to iN conversion

a, Cells immunostained for the neuronal TUBB3 and trophoblast markers CDX2, KRT8 at day 2 after Ascl1 induction. Asterisks indicate cells co-staining for both neuronal and trophoblast markers. **b**, Time resolved immunostaings of cells for neuronal TUBB3 and trophoblast marker KRT8. Asterisks indicate cells co-staining for both neuronal and trophoblast markers. **c**, scRNAseq data generated by Aydin et. al. 2019 showing trophoblast markers *Gata3*, *Krt8* expressed in Ascl1 (labeled as ascl1V5), but not Ngn2 (labeled tubb3gfp), induced cells. **d**, Cells filtered for expression by overexpression cassette (Ascl1 – ascl1V5, Ngn2 – tubb3gfp) or *Gata3*. Ascl1 induced cells co-express both neuronal *Tubb3*, *Dcx2* as well as trophoblast *Gata3* markers.



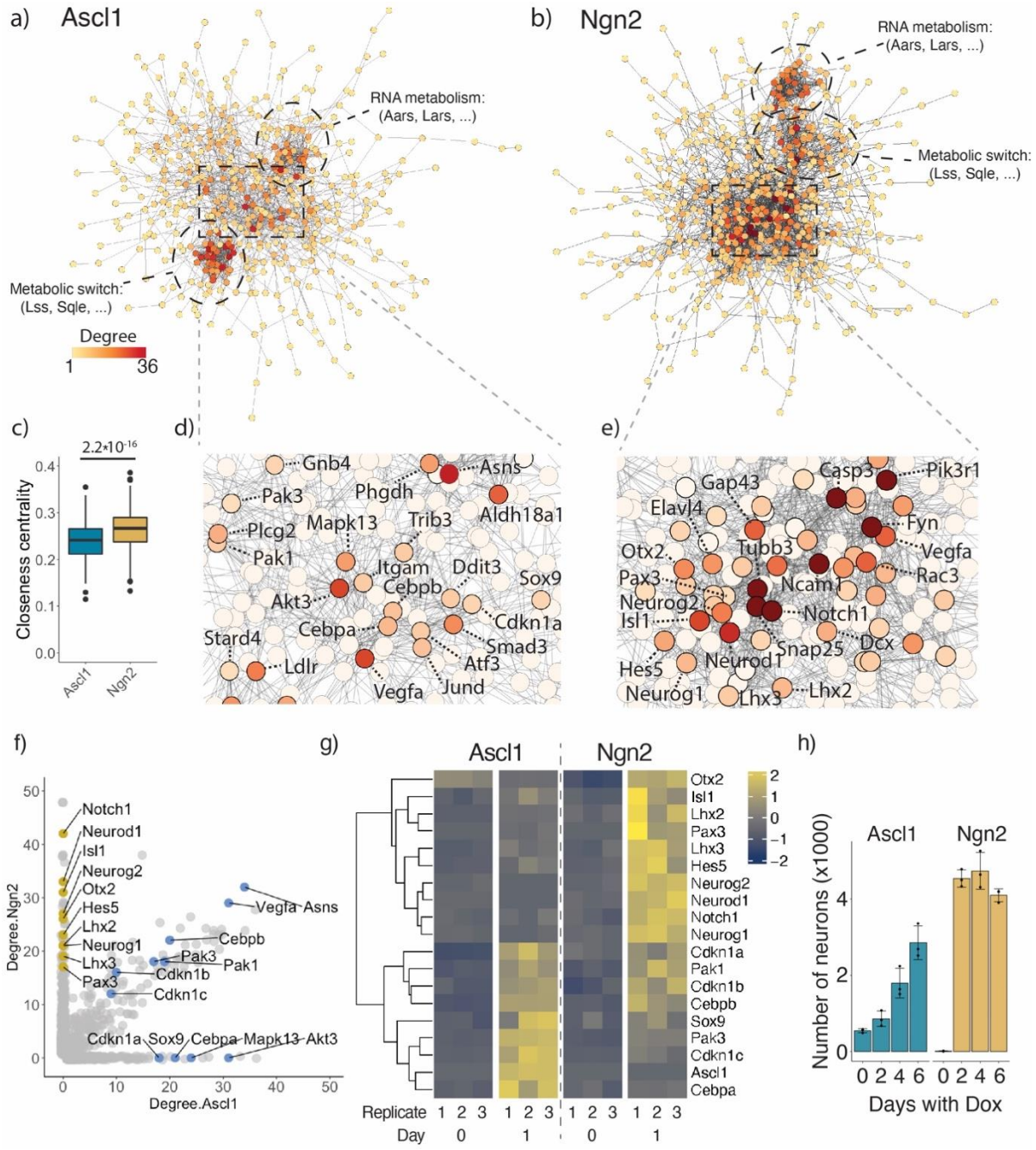
Supplementary Figure 6. Ascl1 and Ngn2 binding analysis

a, ChIPseq heatmap showing Flag-Ascl1 and Flag-Ngn2 binding 1 day after induction. Peaks are grouped based on binding by Ascl1, Ngn2 or both, with the top enriching motifs for the corresponding groups on the right. Peaks are plotted on the heatmap with a ± 1.5kb window around the peak center (n = 4). **b**, Examples of ChIP profiles showing binding of Ascl1 or Ngn2 to alternative lineage genes.



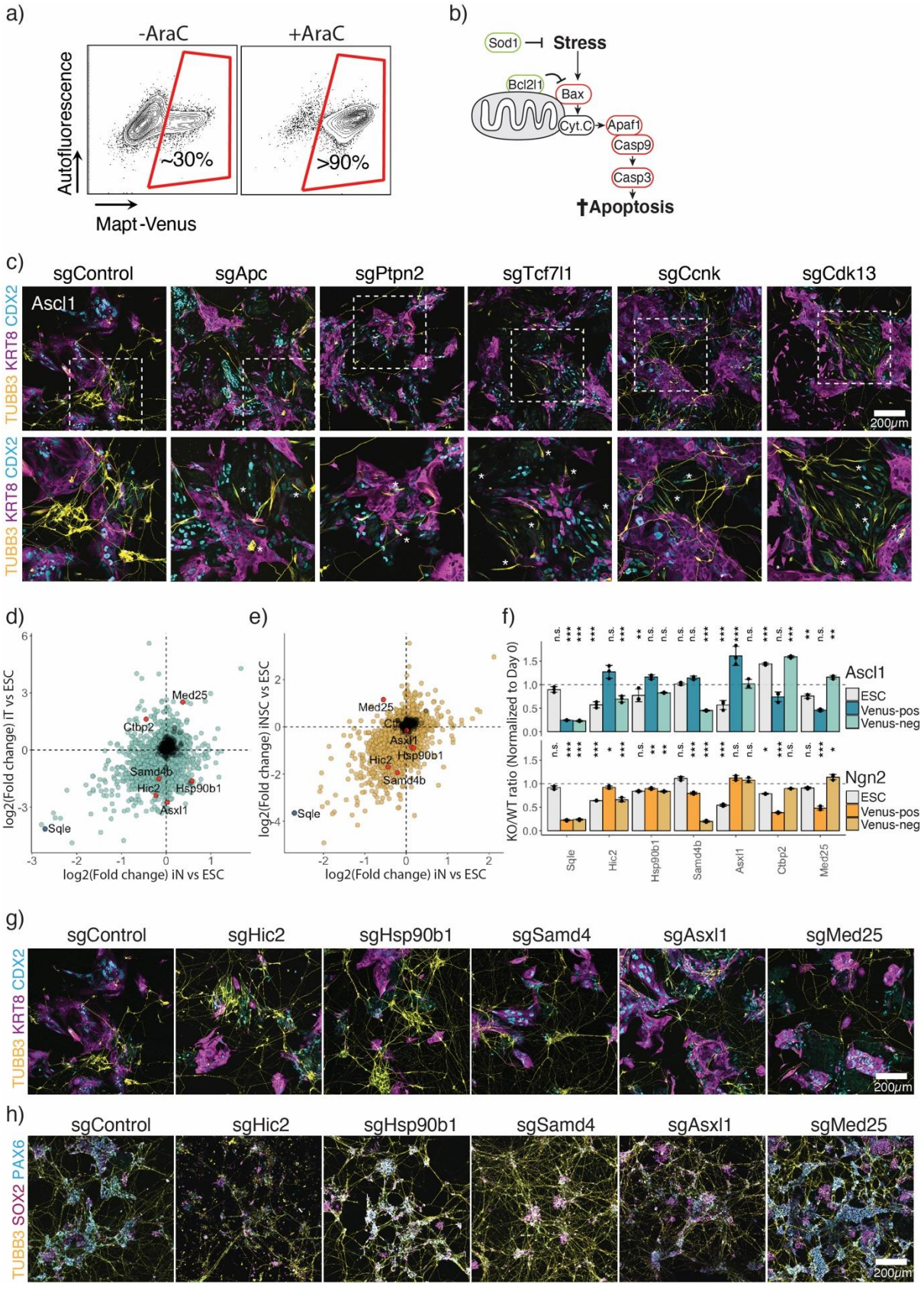
Supplementary Figure 7. Ascl1 binding of alternative lineage markers in ESC and MEF

a, Examples of ChIP profiles showing binding of Ascl1 to the trophoblast lineage markers in ESC and MEF. **b**, Bulk RNAseq expression profiles of the trophoblast lineage markers (**a**) 1 and 2 days post induction of Ascl1 in MEF. Bars show mean of n=2 independent biological replicates. **c**, Examples of ChIP profiles showing Ascl1 or Ngn2 binding of the muscle lineage markers in the ESC. **d**, Expression of muscle lineage markers (**c**) 1 day post induction of Ascl1 or Ngn2 (Fig. 1i).



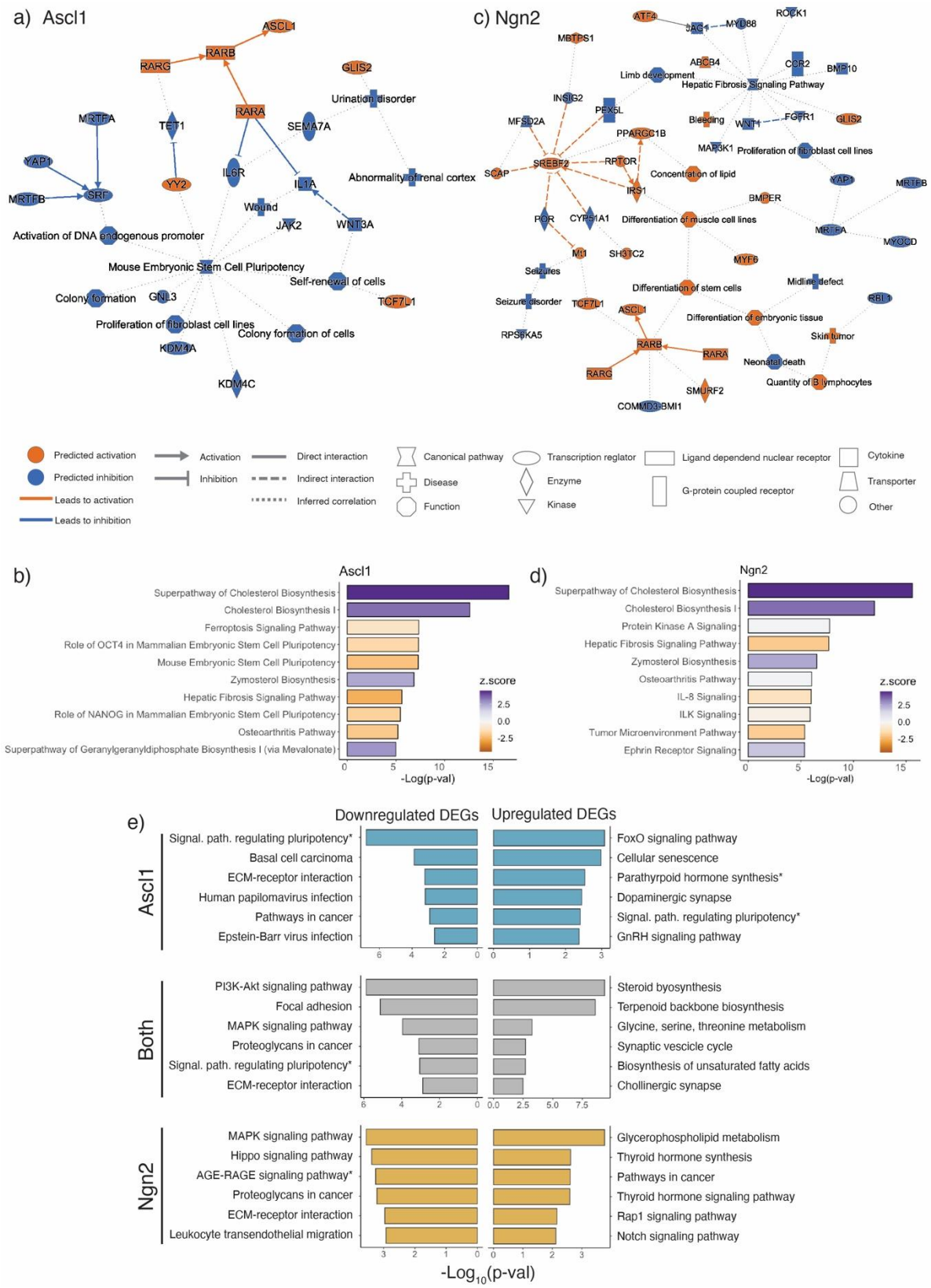
Supplementary Figure 8. Differences in transcriptional responses between Ascl1 or Ngn2 induction

a – b, STRING network reconstruction for Ascl1 (**a**) or Ngn2 (**b**) significantly upregulated genes at day 1 post induction (n = 3, LFC > 1, FDR < 0.05). Values indicate degree of connectivity of a particular gene. Oval dashed line indicates common gene groups between Ascl1 and Ngn2 related to metabolism. Dashed box indicates central nodes in (**d**) and (**e**). **c**, Closeness centrality measure of the (**a**) and (**b**) networks. Boxplots indicate 25th and 75th percentiles as bounds of the box with the median centre line, whiskers indicate minima/maxima of a 1.5x distance of the IQR from the 25th and 75th percentiles, dots indicate outliers outside minima/maxima range. P value of the two-sided Welch two sample t test indicated above. **d – e**, central nodes of the (**a**) and (**b**) networks. Nodes with highest degree of connectivity are labeled. **f**, Scatter plot comparing degree of connectivity between Ascl1 and Ngn2 expressed genes. Central nodes with highest degree of connectivity for Ascl1 are labeled in blue, and highly interconnected neuronal lineage driving genes for Ngn2 are labeled in yellow. **g**, Heatmap showing gene expression for neuronal lineage drivers and example highly interconnected central nodes in Ascl1 expression network (**e**). Color indicates row normalized vst transformed normalized expression of data from Fig. 1i. **h**, Number of neurons as measured by Mapt-Venus reporter after different days of induction of the cassette. Days indicate the period of Dox treatment. Bar plot shows mean of n = 3 biologically independent samples with ± SD.

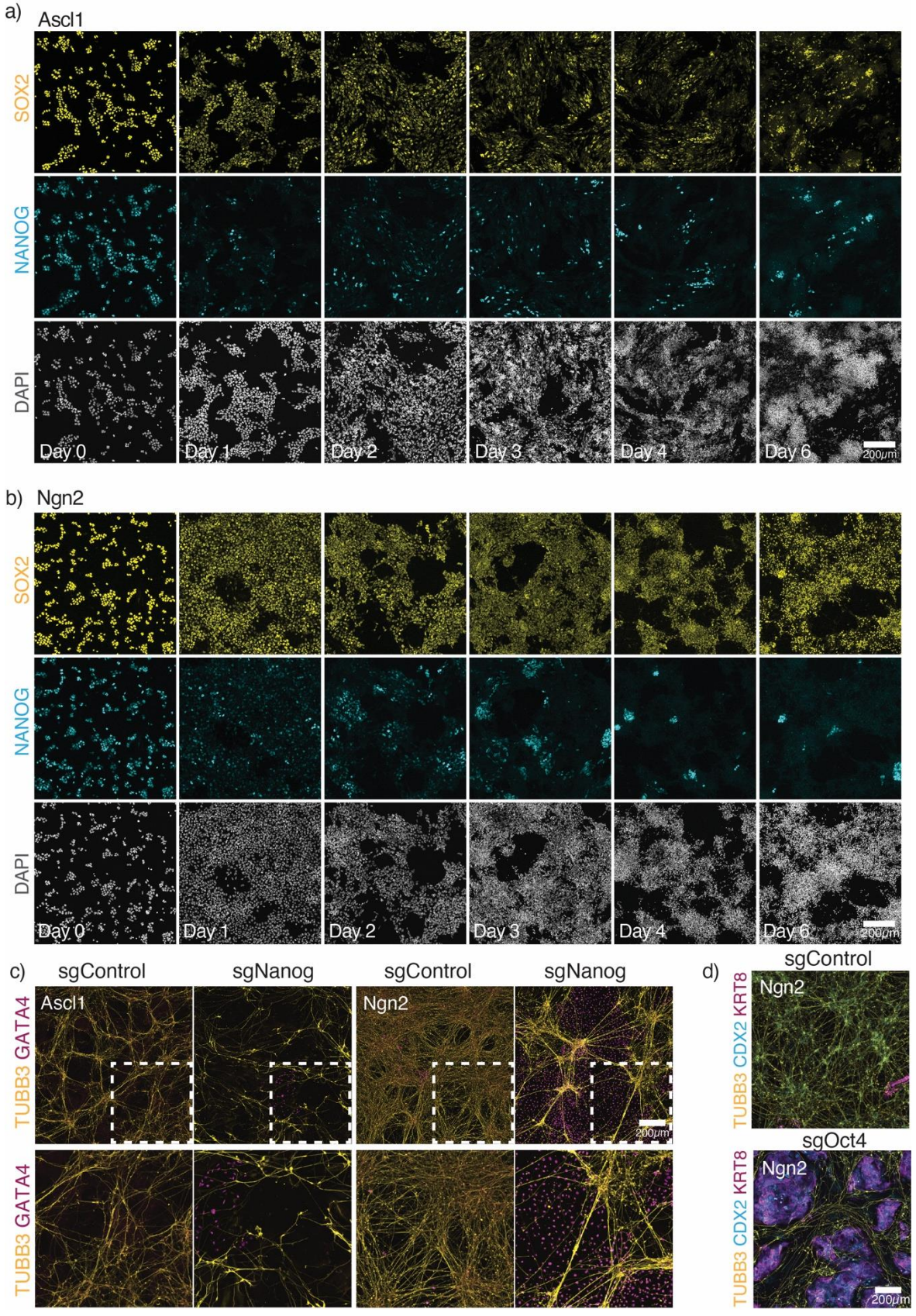

a)
-AraC
+AraC
Autofluorescence
~30%
>90%
Mapt-Venus
b)
Sod1
Stress
Bcl2l1
Bax
Cyt.C
Apaf1
Casp9
Casp3
†Apoptosis
c)
sgControl
sgApc
sgPtpn2
sgTcf7l1
sgCcnk
sgCdk13
Ascl1
TUBB3 KRT8 CDX2
200μm
d)
log2(Fold change) iT vs ESC
log2(Fold change) iN vs ESC
Med25
Ctbp2
Samd4b
Hic2
Hsp90b1
Asxl1
Sqle
e)
log2(Fold change) iNSC vs ESC
log2(Fold change) iN vs ESC
f)
KO/WT ratio (Normalized to Day 0)
Ascl1
Ngn2
ESC
Venus-pos
Venus-neg
Sqle
Hic2
Hsp90b1
Samd4b
Asxl1
Ctbp2
Med25
g)
sgControl
sgHic2
sgHsp90b1
sgSamd4
sgAsxl1
sgMed25
TUBB3 KRT8 CDX2
200μm
h)
sgControl
sgHic2
sgHsp90b1
sgSamd4
sgAsxl1
sgMed25
TUBB3 SOX2 PAX6
200μm

Supplementary Figure 9. Additional CRISPR-Cas9 screen results

a, Purification of Mapt-Venus positive cells by AraC addition at day 4 post induction. Red gate indicating Mapt-Venus positive iN cells. **b**, A reduced scheme depicting apoptosis pathway. Colored proteins indicate hits in the screen: apoptosis promoting and enriching in the screen (red), and apoptosis preventing and depleting in the screen (green). **c**, Immunostaining of the knockout cells for neuronal marker TUBB3 and alternative lineage markers KRT8 / CDX2 for Ascl1 induced iT (Fig. 2h), with TUBB3 positive cells lacking neuronal morphology indicated in asterisks. **D-h**, Additional genes scoring differently in Ascl1 iN vs iT (**d**) and Ngn2 iN vs iNSC, that were validated using FACS (**f**) and immunostaining for neuronal marker Tubb3 (**g-h**) and iT – KRT8, CDX2, for Ascl1 (**g**) and iNSC – SOX2, PAX6, for Ngn2(**h**). Bar plots in **f)** shows mean of n=3 biologically independent samples ± SD. P values, indicated above, were determined by one-way ANOVA followed by Dunnett's multiple comparison test (two-sided) using Ctrl ratio as a control. "n.s." not significant, "*" < 0.05, "**" < 0.01, "***" < 0.001.

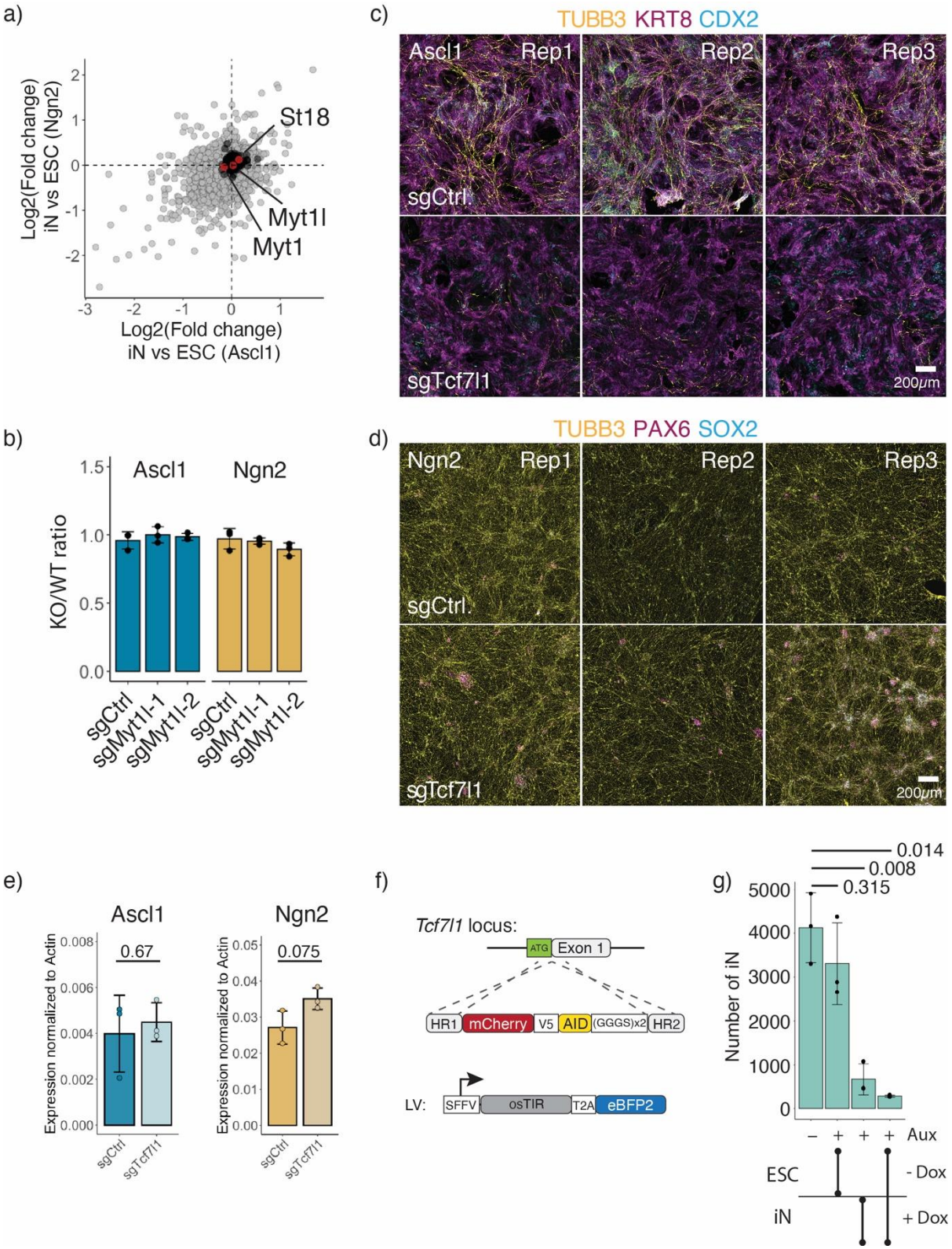

a) Ascl1
ASCL1
RARB
RARG
GLIS2
Urination disorder
RARA
MRTFA
TET1
SEMA7A
YAP1
IL6R
YY2
IL1A
Abnormality of renal cortex
MRTFB
SRF
Wound
JAK2
WNT3A
Activation of DNA endogenous promoter
Mouse Embryonic Stem Cell Pluripotency
Self-renewal of cells
Colony formation
GNL3
TCF7L1
Proliferation of fibroblast cell lines
Colony formation of cells
KDM4A
KDM4C
c) Ngn2
ATF4
SAG1
MYD88
ROCK1
MBTPS1
ABCB4
CCR2
BMP10
Limb development
Hepatic Fibrosis Signaling Pathway
INSIG2
PEX5L
Bleeding
WNT1
FGFR1
GLIS2
MFSD2A
PPARGC1B
MAP3K1
Proliferation of fibroblast cell lines
SCAP
SREBF2
RPTOR
Concentration of lipid
YAP1
MRTFB
IRS1
BMPER
POR
CYP51A1
Differentiation of muscle cell lines
MRTFA
MYOCD
MLL1
SH3TC2
MYF6
Seizures
Differentiation of stem cells
Midline defect
Seizure disorder
TCF7L1
RBL1
ASCL1
Differentiation of embryonic tissue
RPS6KA5
Skin tumor
RARB
RARA
Neonatal death
RARG
Quantity of B lymphocytes
SMURF2
COMMD3-BMI1
Predicted activation
Predicted inhibition
Leads to activation
Leads to inhibition
Activation
Inhibition
Direct interaction
Indirect interaction
Inferred correlation
Canonical pathway
Disease
Function
Transcription reglator
Enzyme
Kinase
Ligand dependend nuclear receptor
G-protein coupled receptor
Cytokine
Transporter
Other
b)
Ascl1
Superpathway of Cholesterol Biosynthesis
Cholesterol Biosynthesis I
Ferroptosis Signaling Pathway
Role of OCT4 in Mammalian Embryonic Stem Cell Pluripotency
Mouse Embryonic Stem Cell Pluripotency
Zymosterol Biosynthesis
Hepatic Fibrosis Signaling Pathway
Role of NANOG in Mammalian Embryonic Stem Cell Pluripotency
Osteoarthritis Pathway
Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via Mevalonate)
z.score
2.5
0.0
-2.5
0
5
10
15
-Log(p-val)
d)
Ngn2
Superpathway of Cholesterol Biosynthesis
Cholesterol Biosynthesis I
Protein Kinase A Signaling
Hepatic Fibrosis Signaling Pathway
Zymosterol Biosynthesis
Osteoarthritis Pathway
IL-8 Signaling
ILK Signaling
Tumor Microenvironment Pathway
Ephrin Receptor Signaling
z.score
2.5
0.0
-2.5
0
5
10
15
-Log(p-val)
e)
Downregulated DEGs
Upregulated DEGs
Ascl1
Signal. path. regulating pluripotency*
Basal cell carcinoma
ECM-receptor interaction
Human papilomavirus infection
Pathways in cancer
Epstein-Barr virus infection
6
4
2
0
0
1
2
3
FoxO signaling pathway
Cellular senescence
Parathyrpoid hormone synthesis*
Dopaminergic synapse
Signal. path. regulating pluripotency*
GnRH signaling pathway
Both
PI3K-Akt signaling pathway
Focal adhesion
MAPK signaling pathway
Proteoglycans in cancer
Signal. path. regulating pluripotency*
ECM-receptor interaction
6
4
2
0
0.0
2.5
5.0
7.5
Steroid byosynthesis
Terpenoid backbone biosynthesis
Glycine, serine, threonine metabolism
Synaptic vescicle cycle
Biosynthesis of unsaturated fatty acids
Chollinergic synapse
Ngn2
MAPK signaling pathway
Hippo signaling pathway
AGE-RAGE signaling pathway*
Proteoglycans in cancer
ECM-receptor interaction
Leukocyte transendothelial migration
3
2
1
0
0
1
2
3
Glycerophospholipid metabolism
Thyroid hormone synthesis
Pathways in cancer
Thyroid hormone signaling pathway
Rap1 signaling pathway
Notch signaling pathway
-Log10(p-val)

Supplementary Figure 10. Additional analysis Ascl1 and Ngn2 overexpression transcriptional response
a, Ingenuity pathway analysis of the differentially expressed (LFC > 1 and LFC < -1, FDR < 0.05) genes 1 day post induction by Ascl1 (Fig. 1i). **b**, Ingenuity canonical pathway enrichment analysis of (**a**). P values were determined using Fisher's exact test. **c**, Ingenuity pathway analysis of the differentially expressed (LFC > 1 and LFC < -1, FDR < 0.05) genes 1 day post induction by Ngn2 (Fig. 1i). **d**, Ingenuity canonical pathway enrichment analysis of (**c**). P values were determined using Fisher's exact test. **e**, Kegg pathway enrichment analysis of the gene subgroups of Fig. 3a. Fisher's exact test was used to calculate p values. Asterisks indicate shortening of terms Signaling pathways regulating pluripotency of stem cells; Parathyroid hormone synthesis, secretion and action; AGE-RAGE signaling pathway in diabetic complications.


a)
Ascl1
SOX2
NANOG
DAPI
Day 0
Day 1
Day 2
Day 3
Day 4
Day 6
200μm
b)
Ngn2
SOX2
NANOG
DAPI
Day 0
Day 1
Day 2
Day 3
Day 4
Day 6
200μm
c)
sgControl
sgNanog
sgControl
sgNanog
TUBB3 GATA4
Ascl1
Ngn2
200μm
TUBB3 GATA4
d)
sgControl
TUBB3 CDX2 KRT8
Ngn2
sgOct4
TUBB3 CDX2 KRT8
Ngn2
200μm

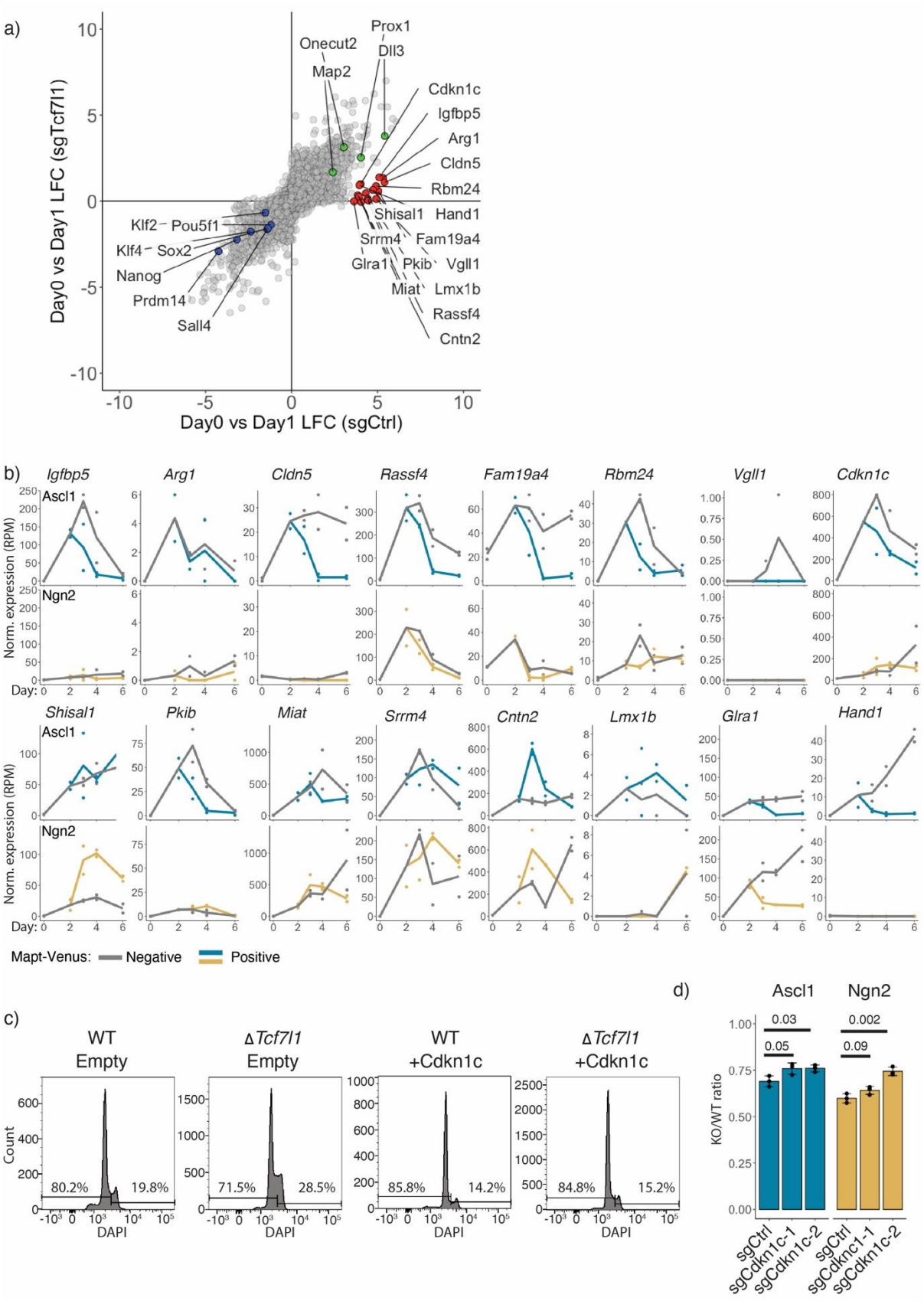
Supplementary Figure 11. Comparison of the pluripotency network shutdown between Ascl1 and Ngn2 induction

a, Time course of the immunostained Ascl1 induced cells for the core pluripotency proteins SOX2 and NANOG. **b**, Time course of the immunostained Ngn2 induced cells for the core pluripotency proteins SOX2 and NANOG. **c**, Immunostaining for the neuronal TUBB3 and primitive endoderm marker GATA4 in the WT or with *Nanog* knockout cells 6 days post Ascl1 or Ngn2 induction. **d**, Immunostaining for the neuronal TUBB3 and trophoblast markers CDX2, KRT8 in the cells with *Pou5f1* (*Oct4)* knockout 6 days post Ngn2 induction.

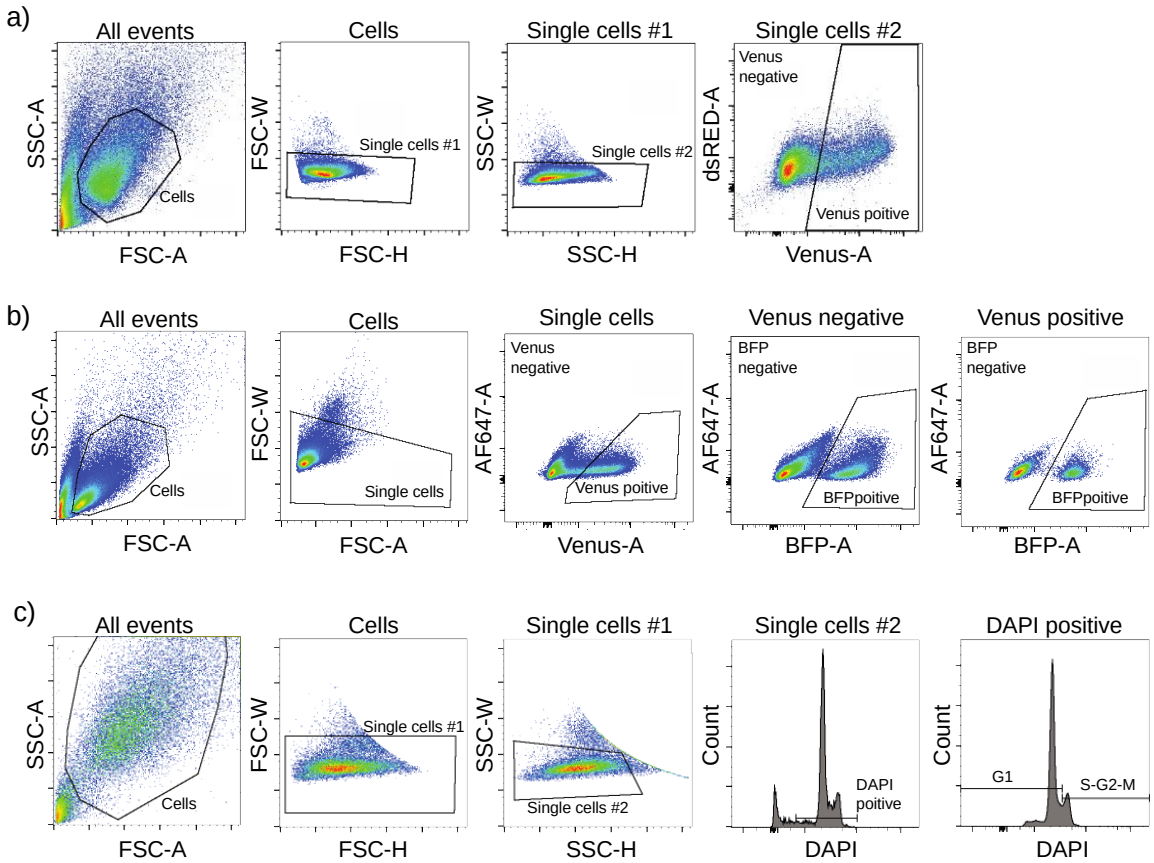

a)
Log2(Fold change) iN vs ESC (Ngn2)
Log2(Fold change) iN vs ESC (Ascl1)
St18
Myt1l
Myt1
b)
Ascl1
Ngn2
KO/WT ratio
sgCtrl
sgMyt1l-1
sgMyt1l-2
c)
TUBB3 KRT8 CDX2
Ascl1
Rep1
Rep2
Rep3
sgCtrl
sgTcf7l1
200μm
d)
TUBB3 PAX6 SOX2
Ngn2
Rep1
Rep2
Rep3
sgCtrl
sgTcf7l1
200μm
e)
Ascl1
Ngn2
Expression normalized to Actin
0.67
0.075
sgCtrl
sgTcf7l1
f)
Tcf7l1 locus:
ATG
Exon 1
HR1
mCherry
V5
AID
(GGGS)x2
HR2
LV:
SFFV
osTIR
T2A
eBFP2
g)
0.014
0.008
0.315
Number of iN
Aux
ESC
iN
- Dox
+ Dox

Supplementary Figure 12. Validation of CRISPR-Cas9 screen results in alternative systems

a, Effects of targeting *Myt1l*, *Myt1*, *St18* genes in CRISPR-Cas9 screen. Axis indicate Log_2(Fold change) of collapsed sgRNA abundance between ESC and iN. **b**, Validation of *Myt1l* knockout effect on iN directed differentiation. Experimental scheme same as Fig. 2f. Bar represents iN efficiency normalized to the efficiency of internal control cell population. Bar plot shows mean of n = 3 biologically independent samples with ± SD. **c-d**, Validation of *Tcf7l1* KO in E14 ESC polyclonally expressing Ascl1 or Ngn2. Immunostainings for neuronal TUBB3 and Trophoblast markers CDX2, KRT8 (Ascl1 (**a**)) or NSC markers PAX6, SOX2 (Ngn2 (**b**)). **e**, Expression of Ascl1 or Ngn2 cassette at day 1 post induction in WT or *Tcf7l1* knockout cells. Bar plot shows mean of n = 3 independent biological replicates with ± SD, p values of the two-sided Welch two sample t test indicated above. **f**, Scheme for N-terminally tagged Tcf7l1 with AID degron and lentiviral expression vector of osTIR. **g,** Efficiency of iN formation of Ascl1 expressing cells with AID-TCF7L1 (**d**) when TCF7L1 is degraded before, after induction or all the time. N = 3 biologically independent samples. Bar plot shows mean with ± SD, p values of the two-sided Welch two sample t test indicated above.


a)
Day0 vs Day1 LFC (sgTcf7l1)
Day0 vs Day1 LFC (sgCtrl)
Onecut2
Prox1
Dll3
Map2
Cdkn1c
Igfbp5
Arg1
Cldn5
Rbm24
Shisal1
Hand1
Srrm4
Fam19a4
Glra1
Pkib
Vgll1
Miat
Lmx1b
Rassf4
Cntn2
Klf2
Pou5f1
Klf4
Sox2
Nanog
Prdm14
Sall4
b)
Igfbp5
Arg1
Cldn5
Rassf4
Fam19a4
Rbm24
Vgll1
Cdkn1c
Shisal1
Pkib
Miat
Srrm4
Cntn2
Lmx1b
Glra1
Hand1
Ascl1
Ngn2
Norm. expression (RPM)
Day:
Mapt-Venus:
Negative
Positive
c)
WT
Empty
ΔTcf7l1
Empty
WT
+Cdkn1c
ΔTcf7l1
+Cdkn1c
Count
80.2%
19.8%
71.5%
28.5%
85.8%
14.2%
84.8%
15.2%
DAPI
d)
Ascl1
Ngn2
KO/WT ratio
0.03
0.05
0.002
0.09
sgCtrl
sgCdkn1c-1
sgCdkn1c-2
sgCtrl
sgCdknc1-1
sgCdkn1c-2

Supplementary Figure 13. Analysis of differentially expressed genes between WT and Tcf7l1 knockout cells.
a, Scatter plot comparing gene expression changes of Ascl1 induction at day 1 post between WT and Tcf7l1 knockout cells. Highlighted circles show genes upregulated in WT and not in *Tcf7l1* KO cells (red), example genes that are neuronal markers (green), core pluripotency genes (blue). N = 3. **b**, Expression over time (Fig. 1a. RNAseq data) of differentially upregulated genes (**(a)** – WT versus *Tcf7l1* KO, red circles). **c**, FACS analysis of cell cycle by DAPI staining at Day 3 between WT and *Tcf7l1* knockouts with or without expression of Cdkn1c. Percentages indicates cells in G1 versus other cell cycle stages. **d**, Effects of *Cdkn1c* knockout on iN formation. Bar represents iN efficiency normalized to the efficiency of internal control cell population. Experiment performed as a competition assay like in Fig. 2f. Normalized efficiency = (% of neurons in KO / % of neurons in WT). N = 3 biologically independent samples. Bar plot shows mean with ± SD, p values of the two-sided Welch two sample t test indicated above.



Supplementary Figure 14. FACS gating strategies

a, Representative FACS gating strategy to assess *Mapt*-Venus reporter in Fig. 3g-i, 4f, Supplement Fig. 2j, 4e, 8h, 9a, 12g. dsRED channel was used for assessing autofluorescence. **b**, Gating strategy used in competition assays in Fig. 2g, Supplement Fig. 9f, 12b, 13d. Alexa Fluor 647 (AF647) channel was used for measuring autofluorescence. **c,** Gating strategy for the supplement fig. 13c. Signal peak parameters are as follows: A – area, H – height, W – width.

Supplementary Table 1. Primers used for the sgRNA cloning.

sgRNA	Forward	Reverse
Ctrl	CACCGGAGTCGTTTTACCCGCCGC	AAACGCGGCGGGTAAAACGACTCC
Tcf7l1	CACCGCCTTCGGCGAAATAGTCGCG	AAACCGCGACTATTTCGCCGAAGGC
Sqle	CACCGAGTGTCGAATCAACACCAGA	AAACTCTGGTGTTGATTCGACACTC
Apc	CACCGGTACACCTGCTGAATACGAG	AAACCTCGTATTCAGCAGGTGTACC
Ptpn2	CACCGAAGAAGTTACATCTTAACAC	AAACGTGTTAAGATGTAACTTCTTC
Strap	CACCGGGCAAGCCCATGCTCCGCCA	AAACTGGCGGAGCATGGGCTTGCCC
Cnot8	CACCGTGTAGCTGAGGACAATCTCA	AAACTGAGATTGTCCTCAGCTACAC
Ccnk	CACCGTACCTTCCAGAACGTCAACA	AAACTGTTGACGTTCTGGAAGGTAC
Cdk13	CACCGATAAGTCGAGTGCAGCTGCG	AAACCGCAGCTGCACTCGACTTATC
Setd1b	CACCGCATGGGCAACATTATCCACG	AAACCGTGGATAATGTTGCCCATGC
Nanog_1	CACCGTCTGAACCTGAGCTATAAGC	AAACGCTTATAGCTCAGGTTCAGAC
Nanog_2	CACCGTAAGCAAGAATAGTTCTC	AAACGAGAACTATTCTTGCTTAC
Pou5f1_1	CACCGATCACCTTGGGGTACACCC	AAACGGGTGTACCCCAAGGTGATC
Pou5f1_2	CACCGCCCAGGGTGAGCCCCACGT	AAACACGTGGGGCTCACCCTGGGC
Sox2_1	CACCGCAGGGCGCTGACGTCGTAG	AAACCTACGACGTCAGCGCCCTGC
Sox2_2	CACCGAGCCCAGCGCCATACCGG	AAACCCGGTATGGCGCTGGGCTC
Cdkn1c_1	CACCGCTACGCGCTATCACTGGGA	AAACTCCCAGTGATAGCGCGTAGC
Cdkn1c_2	CACCGACAGCCACGGCCACCGG	AAACCCGGTGGCCGTGGCTGTC
Myt1l_1	CACCGATGGGTCAGGACACGTCAG	AAACCTGACGTGTCCTGACCCATC
Myt1l_2	CACCGAAGAAAGACGGTATCCAG	AAACCTGGATACCGTCTTTCTTC
Mapt_C-terminus_tag	CACCGGCCAAGCAGGGTTTGTGATC	AAACGATCACAAACCCTGCTTGGCC
Tcf7l1_N-terminus_tag	CACCGCCACCGAGCTGGGGCATGGT	AAACACCATGCCCCAGCTCGGTGGC

Supplementary Table 2. List of antibodies used.

Antibody	Host species	Producer	Catalogue #	Clone number	IF dilution
anti-Tubb3	Mouse	Sigma	T8660	SDL.3D10	1:500
anti-Tubb3	Rabbit	Biolegend/Covance	PRB-435P	-	1:500
anti-Map2	Rabbit	Abcam	ab32454	-	1:500
anti-Sox2	Rat	Invitrogen (eBioscience)	14-9811-80	Btjce	1:500
anti-Oct4	Rabbit	Abcam	ab19857	-	1:500
anti-Nanog	Rabbit	Abcam	ab80892	-	1:500
anti-Pax6	Rabbit	Covance	PRB-278P	-	1:200
anti-Krt8	Rat	DSHB	AB 531826	-	1:200
anti-Cdx2	Rabbit	Abcam	ab76541	EPR2764Y	1:400
anti-Nestin	Mouse	Merk	MAB353	rat-401	1:200
anti-Gata4	Rat	Invitrogen	14998082	eBioEvan	1:500
anti-Cdkn1c	Rabbit	Abcam	ab75974	-	1:250
anti-Mki67	Rat	Invitrogen (eBioscience)	14-5698-82	SolA15	1:200
anti-Ascl1	Mouse	Invitrogen (eBioscience)	14-5794-82	24B72D11	1:200
anti-Tpbpa	Rabbit	Abcam	ab104401		1:200
anti-Flag	Mouse	Sigma	F1804	M2	5ug/ChIP sample
anti-Mouse-488	Goat	Invitrogen	A11029	-	1:1000
anti-Rabbit-488	Goat	Invitrogen	A11034	-	1:1000
anti-Rat-488	Goat	Invitrogen	A11006	-	1:1000
anti-Mouse-568	Goat	Invitrogen	A11031	-	1:1000
anti-Rabbit-568	Goat	Invitrogen	A11036	-	1:1000
anti-Mouse-647	Goat	Invitrogen	A21247	-	1:1000
anti-Rabbit-647	Goat	Invitrogen	A21236	-	1:1000
anti-Rat-647	Goat	Invitrogen	A21245	-	1:1000

Supplementary Table 3. qPCR primers used.

Gene	Forward	Reverse
Nanog	GCCTCCAGCAGATGCAAGAAC	CTGGTGCTGAGCCCTTCTG
Pou5f1	GAGTCCCAGGACATGAAAGCC	CAGATGGTGGTCTGGCTGAAC
Sox2	CATGAGAGCAAGTACTGGCAAG	CCAACGATATCAACCTGCATGG
Acta	GCTGTATTCCCCTCCATCGTG	CACGGTTGGCCTTAGGGTTCAG
Ngn2-Mwcassette	GGCCCCGAATTCGCCACCAT	AGCTCCTCGTCCTCCTCCTC
Ascl1-Mwcassete	CCGAATTCGCTAGCCACCAT	AAGAAGCAGGCTGCGGG

CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens

Georg Michlits[1], Maria Hubmann[1], Szu-Hsien Wu[1], Gintautas Vainorius[1], Elena Budusan[1], Sergei Zhuk[1], Thomas R Burkard[1,2], Maria Novatchkova[1,2], Martin Aichinger[2], Yiqing Lu[3,4], John Reece-Hoyes[5], Roberto Nitsch[6], Daniel Schramek[3,4], Dominic Hoepfner[7] & Ulrich Elling[1]

Pooled CRISPR screens are a powerful tool for assessments of gene function. However, conventional analysis is based exclusively on the relative abundance of integrated single guide RNAs (sgRNAs) between populations, which does not discern distinct phenotypes and editing outcomes generated by identical sgRNAs. Here we present CRISPR-UMI, a single-cell lineage-tracing methodology for pooled screening to account for cell heterogeneity. We generated complex sgRNA libraries with unique molecular identifiers (UMIs) that allowed for screening of clonally expanded, individually tagged cells. A proof-of-principle CRISPR-UMI negative-selection screen provided increased sensitivity and robustness compared with conventional analysis by accounting for underlying cellular and editing-outcome heterogeneity and detection of outlier clones. Furthermore, a CRISPR-UMI positive-selection screen uncovered new roadblocks in reprogramming mouse embryonic fibroblasts as pluripotent stem cells, distinguishing reprogramming frequency and speed (i.e., effect size and probability). CRISPR-UMI boosts the predictive power, sensitivity, and information content of pooled CRISPR screens.

CRISPR–Cas9 is frequently used to disrupt gene function in pooled genetic screens[1–5]. To generate a pool of mutants, researchers transduce cells expressing the Cas9 endonuclease with viral libraries that encode sgRNAs. The Cas9–sgRNA complex generates double-strand breaks (DSBs) at target genes, creating diverse mutations via error-prone nonhomologous end-joining (NHEJ). Currently, the gene linked to a given phenotype is inferred from the relative abundance of integrated sgRNAs in populations before and after selection. This approach does not account for distinct mutations generated by CRISPR–Cas9. Although out-of-frame deletions/insertions generate the desired loss-of-function (LOF) alleles, in-frame deletions/insertions may generate hypermorphic, hypomorphic, or neutral alleles that survive dropout screens, thus limiting the maximal achievable depletion. The probability of biallelic deleterious gene mutations is estimated to be 30–70% (refs. 6–8). Together with stochastic noise in the screening process, underlying cell heterogeneity, and biases inherent to subsequent PCR and high-throughput sequencing, clonal effects and conceptual depletion limits can interfere with the acuity of screen readouts[8,9].

Here we report CRISPR-UMI, a UMI-based pooled CRISPR screening approach that can be used to analyze the outcome of genetic screens at the single-cell level. CRISPR-UMI overcomes the current limitations of screen analyses, improves hit identification and quantification, and enables phenotypic assessment at the single-cell level.

RESULTS

A framework for single-cell-based CRISPR screening

Conceptually, the depletion limit in CRISPR screens applies to analyses of populations (**Fig. 1a**). Evaluation of independently derived single-cell clones as biological replicates in a pooled CRISPR screen overcomes the depletion limit by discerning clones with homozygous LOF alleles that are depleted from clones with other (non-LOF) alleles that survive (**Fig. 1b**). Importantly, the response of LOF clones to selection is a true reflection of biological effect and is independent of editing efficiency. This binary mode of scoring can be leveraged if (i) cells infected with an sgRNA can be individually tracked by an UMI, and (ii) the cell population is carried through a strong bottleneck so that a maximum of one cell and thus one editing outcome per UMI remains in the population (**Supplementary Fig. 1a**,**b**). This method renders screens more informative by distinguishing the strong selection of a few clones from the mild selection of many clones, in both negative and positive selection (**Fig. 1b**,**c**).

Generating a complex CRISPR library with random barcodes

To ensure optimal sgRNA efficiency, minimal off-targeting, and a likely biological effect on-target, we predicted Doench

[1]Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Vienna, Austria. [2]Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC),Vienna, Austria. [3]Center for Systems Biology, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [4]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [5]Novartis Institutes for BioMedical Research, Cambridge, Massachusetts, USA. [6]Discovery Sciences RAD, AstraZeneca R&D, Gothenburg, Sweden. [7]Novartis Institutes for BioMedical Research, Basel, Switzerland. Correspondence should be addressed to U.E. (ulrich.elling@imba.oeaw.ac.at).

RECEIVED 5 APRIL; ACCEPTED 11 SEPTEMBER; PUBLISHED ONLINE 16 OCTOBER 2017; DOI:10.1038/NMETH.4466

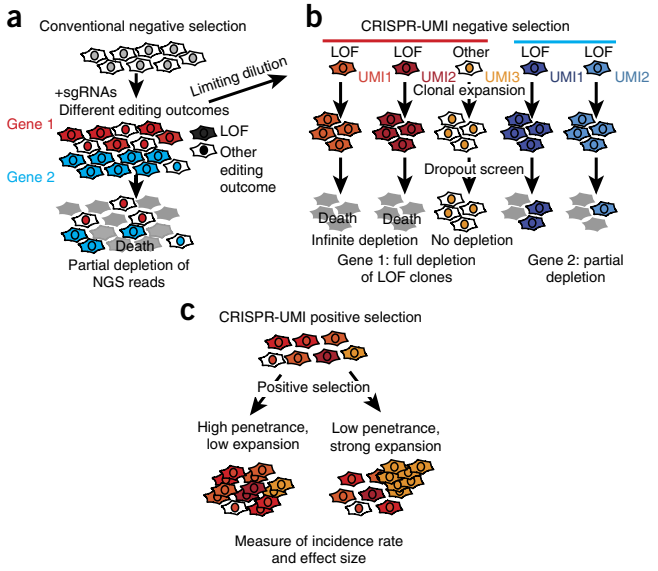


Figure 1 | The conceptual framework of CRISPR screen analysis by single-cell tracing. (**a**,**b**) Schemes describing (**a**) conventional negative-selection (dropout) screens (**a**) compared with CRISPR-UMI (**b**). (**c**) CRISPR-UMI positive-selection screens differentiate incidents and effect size.

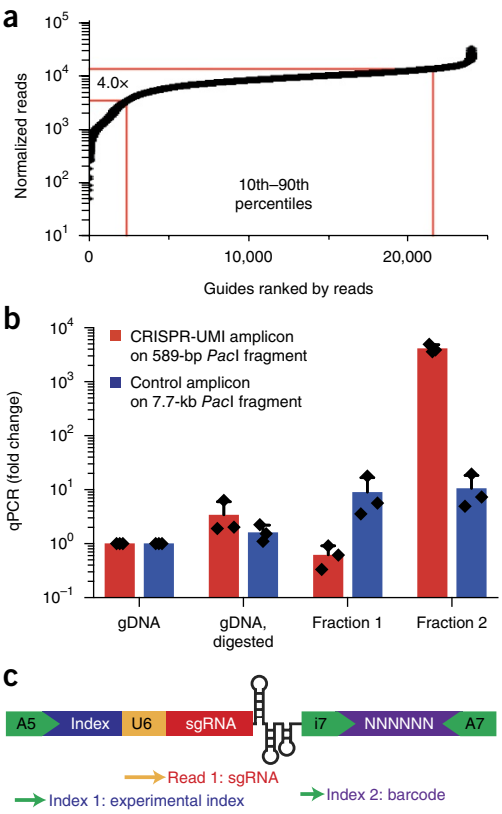


Figure 2 | CRISPR-UMI library representation and post-screen enrichment. (**a**) Guide representation in sgRNA libraries with reads normalized between individual subpools. (**b**) qPCR of genomic DNA (gDNA), gDNA after *Pac*I digestion, and fractions 1 and 2 containing large and small DNA fragments, respectively. Error bars indicate s.d. The results shown are from one experiment conducted in technical triplicate representative of three experiments. (**c**) The barcode-tagged sgRNA design for CRISPR-UMI and the next-generation sequencing strategy. Illumina primer binding and adaptor sequences are shown in green.

scores[7] for all possible sgRNAs in the mouse genome. We subsequently considered off-target predictions, position within the on-target transcripts, exon structure, and protein-coding domains (**Supplementary Fig. 2**). CRISPR-UMI requires the generation of a high-complexity library for clonal tracking of individual cells. We introduced a barcode of ten random nucleotides by parallel cloning into a retroviral backbone. Subsequently, we cloned several subpools of sgRNAs targeting 6,560 different mouse genes, including all genes that encode nuclear proteins, with 4 sgRNAs per gene and 112 nontargeting controls[3], into the vector-barcode library at a coverage of 954–8,776 independent cloning events per sgRNA. This strategy combined each sgRNA with a different barcode in each ligation event. Each sgRNA–barcode pair represented a UMI (**Supplementary Figs. 3** and **4a**,**b**, **Supplementary Table 1**). The relative difference in the abundance of guides in the library at the 10th and 90th percentiles was fourfold (**Fig. 2a**). The overall library complexity reached 83.5 million, which exceeds the number of clones assayed in a screen. Genomic-DNA-containing sgRNA integrations could be enriched 10^3- to 10^4-fold before PCR amplification by *Pac*I digestion and subsequent size selection (**Fig. 2b**, **Supplementary Fig. 4c**,**d**). This enrichment minimized the number of PCR cycles required for amplification, thereby reducing PCR amplification biases. Our vector design enabled direct sequencing of the PCR product (**Fig. 2c**). Thus, our libraries targeted mostly protein-coding domains of nuclear proteins, with even distribution and highly complex UMIs, allowing for single-cell-based CRISPR screens.

Single-cell lineage tracing improves signal-to-noise ratio in dropout CRISPR screening

To test CRISPR-UMI, we conducted a dropout screen by exposing cells to etoposide, which generates DSBs via inhibition of topoisomerase II (ref. 10). Cells defective in DSB repair, such as those with compromised NHEJ, are expected to be sensitive to selection[11].

To optimize the conditions for CRISPR-UMI, we carried out a pilot etoposide screen on 365 genes (1,437 guides) associated with DNA damage. We infected mouse embryonic stem cells (mESCs) harboring doxycycline (Dox)-inducible Cas9 with retroviral vectors delivering sgRNAs and carried out a limiting dilution and expansion to generate a matrix of 4,000 cells per guide in four conditions: (i) 20 clones of 200 cells, (ii) 64 clones of 64 cells, (iii) 200 clones of 20 cells, and (iv) no limiting dilution, resulting in 4,000 mostly independently infected cells (**Supplementary Fig. 5a**–**c**). We evaluated the performance of sgRNAs that targeted core members of the NHEJ pathway. For every sgRNA, we determined the median depletion of individual clones (**Supplementary Fig. 5d**) and calculated *P* values, combining independent clones that carried the same sgRNAs by using model-based analysis of genome-wide CRISPR–Cas9 knockout (MAGeCK)[12], similarly to combining independent sgRNAs targeting the same gene. Thus, we analyzed population behavior at a clonal level. We identified all core members of the NHEJ complex and found that multiple guides had highly significant scores ($P < 10^{-10}$), which validated the quality of the guide prediction and library. The performance of NHEJ-pathway members relative to control sgRNAs, as measured by the signal-to-noise ratio (SNR) (**Supplementary Fig. 5e**),

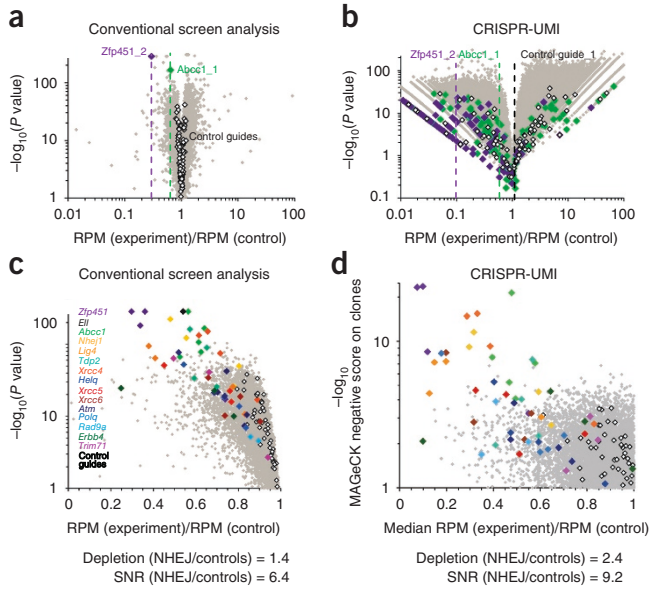


Figure 3 | Conventional and clonal analysis of negative-selection screens. (**a**,**b**) Volcano plots of (**a**) conventional CRISPR analysis and (**b**) single-cell-derived clones (CRISPR-UMI analysis). The median depletions of two guides are shown as dashed lines. In both plots, the *x*-axis shows sgRNA representation relative to the control, and the *y*-axis shows the binomial *P* value. (**c**) Conventional analysis of depleting sgRNAs; axes are as in **a**,**b**. (**d**) CRISPR-UMI analysis of the same sgRNAs as in **c** (color-coded as in **c**). *P* values are based on MAGeCK scores of clones; the *x*-axis indicates the median depletion of clones. The depletion level and signal-to-noise ratio are shown below. RPM, reads per million mapped reads.

reached similar values of 15–18 in all conditions but was best at 148 clones of 35 reads. On the basis of the sgRNA distribution in our library, this condition also results in the lowest number of guides represented in too few clones for analysis (0.06% versus 1.4% or 6.6%). We therefore performed single-cell lineage-tracing CRISPR-UMI screening at roughly 150 clones with an average of 30–40 reads.

Using this optimal clone number and size, we tested CRISPR-UMI in an etoposide screen with the full library of 26,514 guides targeting 6,560 genes (**Supplementary Fig. 6**). We compared performance at the guide level in CRISPR-based negative-selection screens and observed clonal variance, as well as increased levels of median depletion at the clonal level (**Fig. 3a**,**b**). To determine whether CRISPR-UMI outperforms conventional analysis, we combined all clones per guide and plotted the median depletion for each guide versus a *P* value computed by MAGeCK[12] (**Fig. 3c**,**d**). Indeed, CRISPR-UMI improved the SNR ($SNR_{CRISPR\text{-}UMI}$ = 9.2 versus SNR_{CRISPR} = 6.4 for NHEJ versus control) and depletion ($Depletion_{CRISPR\text{-}UMI}$ = 2.4 versus $Depletion_{CRISPR}$ = 1.4 for NHEJ versus control) compared with conventional analysis.

To directly compare the performance of conventional versus CRISPR-UMI analysis, we ranked guides in each data set, generating a guide score. We observed a high degree of correlation between conventional and single-cell-based screen analysis, with some guides uncovered by only one approach (**Fig. 4a**). Intriguingly, all discrepancies correlated with outlier clones having overrepresented read numbers that dominated the total read space (**Fig. 4b**, **Supplementary Fig. 7**). These outliers interfered with conventional analysis, which assumes that cells carrying the same sgRNA are equally represented. For guides Trim71_1 and Ell_2, scoring in conventional analysis, outlier clones showed depletion, masking the fact that all other clones showed no obvious tendency for depletion. In contrast, outlier clones for CRISPR-UMI-specific hits such as Lig4_3 and Rad9a_4 showed enrichment, masking the depletion of most other clones.

To test whether these outlier clones caused the discrepancies, we reanalyzed the data after removing the two clones with the most reads. This improved the alignment of the two analysis pipelines (**Fig. 4a**). After removal of the outliers, guides previously identified only by CRISPR-UMI were also uncovered by conventional analysis, whereas the ranking of guides previously identified only by conventional analysis decreased (Trim71_1, from 19 to 398; Ell_2, from 7 to 24,587), as reflected in their lower guide scores (Trim71_1, from 5.8 to 3.4; Ell_2, from 6.5 to 0.07) in the conventional analysis. This also resulted in a considerable drop in the MAGeCK rankings on the gene level (*Trim71*, from 9 to 214; *Ell*, from 20 to 4,416). Thus, the guides identified solely by population-based methods were false positives, and those missed by conventional methods were false negatives. We conclude that undetected outlier clones confound conventional but not CRISPR-UMI analyses. We repeated the etoposide screen, omitting the limiting dilution step, and again found outlier clones differentially called by conventional analysis versus CRISPR-UMI, thus demonstrating that outlier clones are not introduced by limiting dilution (**Supplementary Fig. 8a**–**c**). CRISPR-UMI analysis of pooled screens is thus superior to conventional analysis, because it avoids artifacts that arise from clonal variation and outliers.

To compare the performance of the two scoring methods directly, we overlapped the top 50 and 100 sgRNAs with the NHEJ-pathway genes (*Lig4*, *Xrcc4*, *Xrcc5*, *Xrcc6*, *Nhej1*) as positive controls (**Supplementary Fig. 9**). Whereas conventional analysis called only 7 and 8 out of 21 sgRNAs, CRISPR-UMI scored 12 and 13 sgRNAs within the top 50 and 100 sgRNAs, respectively. Additionally, we determined the reproducibility on the sgRNA level of all genes identified in the screen. We plotted the average number of sgRNAs present per gene (for all genes hit by the respective group of guides) as a function of increasing numbers of sgRNAs sorted by rank (**Fig. 4c**; for example, if 15 genes hit within the top 30 sgRNAs, the value was 2; a value of 1 would be expected for a random data set). For both screens (with and without clonal dilution/expansion) CRISPR-UMI outperformed conventional analysis across the entire hit list and showed higher reproducibility between sgRNAs.

For comparison at the gene level, we combined guides by using MAGeCK for conventional analysis and Fisher's method for CRISPR-UMI and ranked the guides according to score for each method (**Fig. 4d**). Both methods scored multiple expected hits in DNA-repair pathways[11,13,14]. Furthermore, both methods identified the ABC-transporter-encoding gene *Abcc1*, as well as *Zfp451*, which encodes a SUMO E3 ligase. CRISPR-UMI outperformed conventional analysis methods not only on the guide level but also on the gene level, showing stronger depletion and better ranking of hits. To test whether the genes identified in the negative-selection screen could be experimentally validated, we derived knockout cell lines for common (*Lig4* and *Zfp451*), conventional-analysis-specific (*Slc25a4*, *Adcy3*, *Trim71*), and CRISPR-UMI-specific hits (*Rad9a*, *Erbb4*, *Rac1*) and quantified

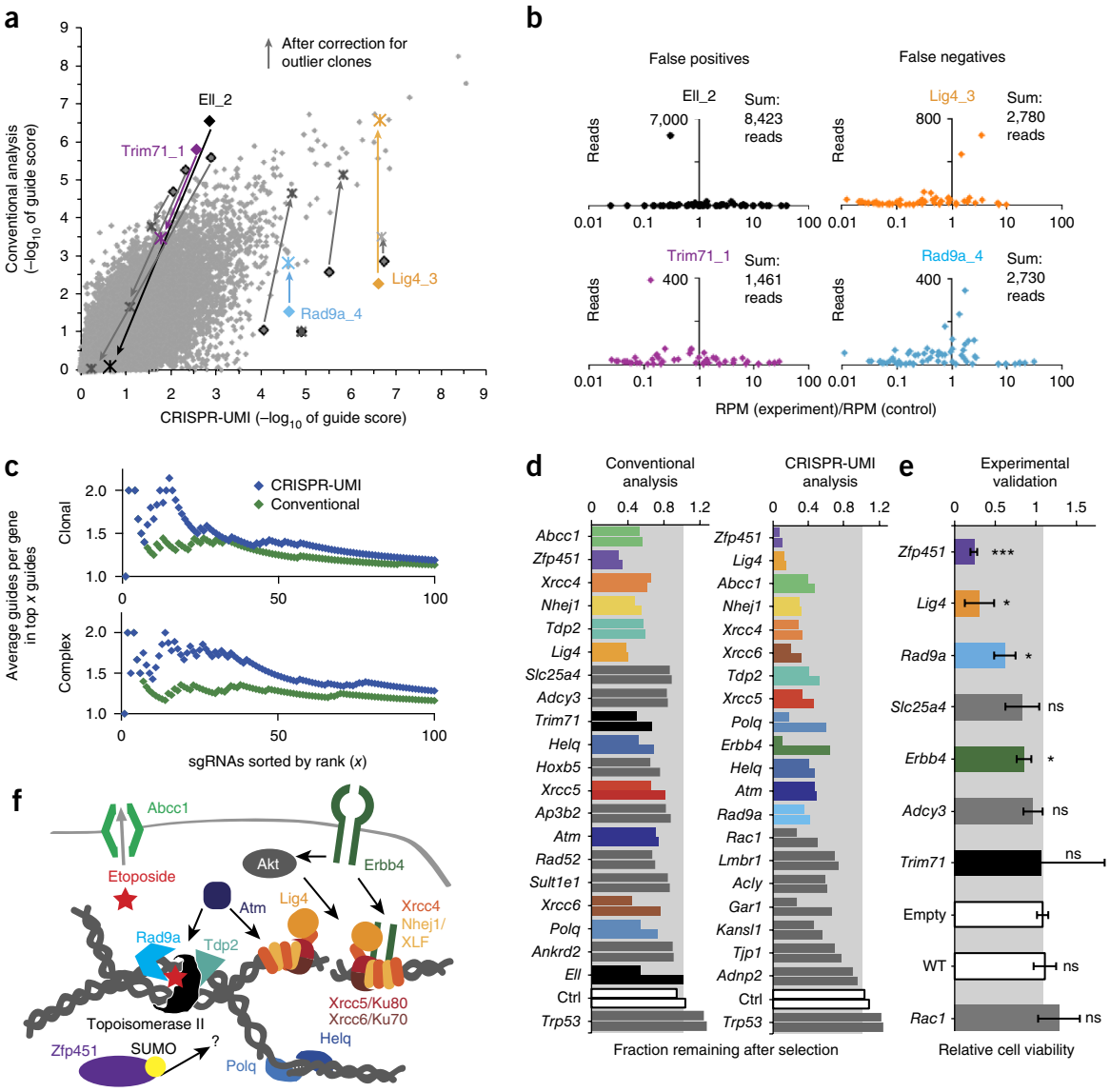


Figure 4 | The performance of conventional analysis compared with that of CRISPR-UMI. (**a**) A comparison of conventional and single-cell-based analysis at the guide level; discrepant guides are highlighted before (diamonds) and after (asterisks) outlier removal. (**b**) Plots for discrepant guides in individual clones showing abundance relative to untreated controls (*x*-axis) versus the total number of reads per clone (*y*-axis). RPM, reads per million mapped reads. (**c**) The average number of guides targeting the same gene (*y*-axis) for genes correlated with the top sgRNAs (*x*-axis), sorted by rank. (**d**) The top-ranking genes as determined by conventional analysis and CRISPR-UMI; depletion of the two best guides per gene is shown. Genes are color-coded according to function. (**e**) Validation of *Lig4*, *Zfp451*, *Rad9a*, and *Erbb4* showing sensitization to etoposide. Four clones per gene (2 for *Trim71* and *Rac1*); three technical replicates each. WT, wild type. Data shown represent the median ± s.e.m.; a heteroskedastic, two-sided *t*-test was applied. ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$; ns, $P > 0.05$. (**f**) A visual summary of the proteins encoded by all genes identified by CRISPR-UMI. Depicted are proposed functions in etoposide removal, topoisomerase II–DNA conjugate clearance, and DNA double-strand break signaling and repair.

dropout in response to etoposide relative to that in the control (**Fig. 4e**, **Supplementary Table 2**). Common hits showed strong dropout; *Rad9a* and *Erbb4* also were depleted significantly (*Rad9a*, $P = 0.012$; *Erbb4*, $P = 0.039$) and with correlating effect size, whereas the conventional-analysis-specific hits were not validated. Genes identified by CRISPR-UMI, namely, *Rad9a* and *Erbb4* (**Fig. 4e**, **Supplementary Fig. 7c**), were both previously implicated in DSB repair or decatenation of DNA in response to topoisomerase II inhibition[15–20]. Taken together, our results identified multiple known proteins, as well as novel proteins, involved in DNA repair after etoposide-mediated topoisomerase II inhibition (**Fig. 4f**).

Positive-selection screen to elucidate roadblocks to reprogramming

Next, we chose a robust induced pluripotent stem cell (iPSC) induction protocol[21] to test CRISPR-UMI in a positive-selection paradigm. We collected mouse embryonic fibroblasts (MEFs) that contained a Dox-inducible *Oct4* (*Pou5f1*)-*Klf4*-*Sox2*-*Myc* (OKSM) cassette, as well as an endogenous Oct4–GFP reporter to detect successful reprogramming[21]. We infected these MEFs with a lentiviral construct that delivered Cas9 and, subsequently, our CRISPR-UMI library (**Fig. 5a**). Six days post-infection (day 0), we induced OKSM by Dox

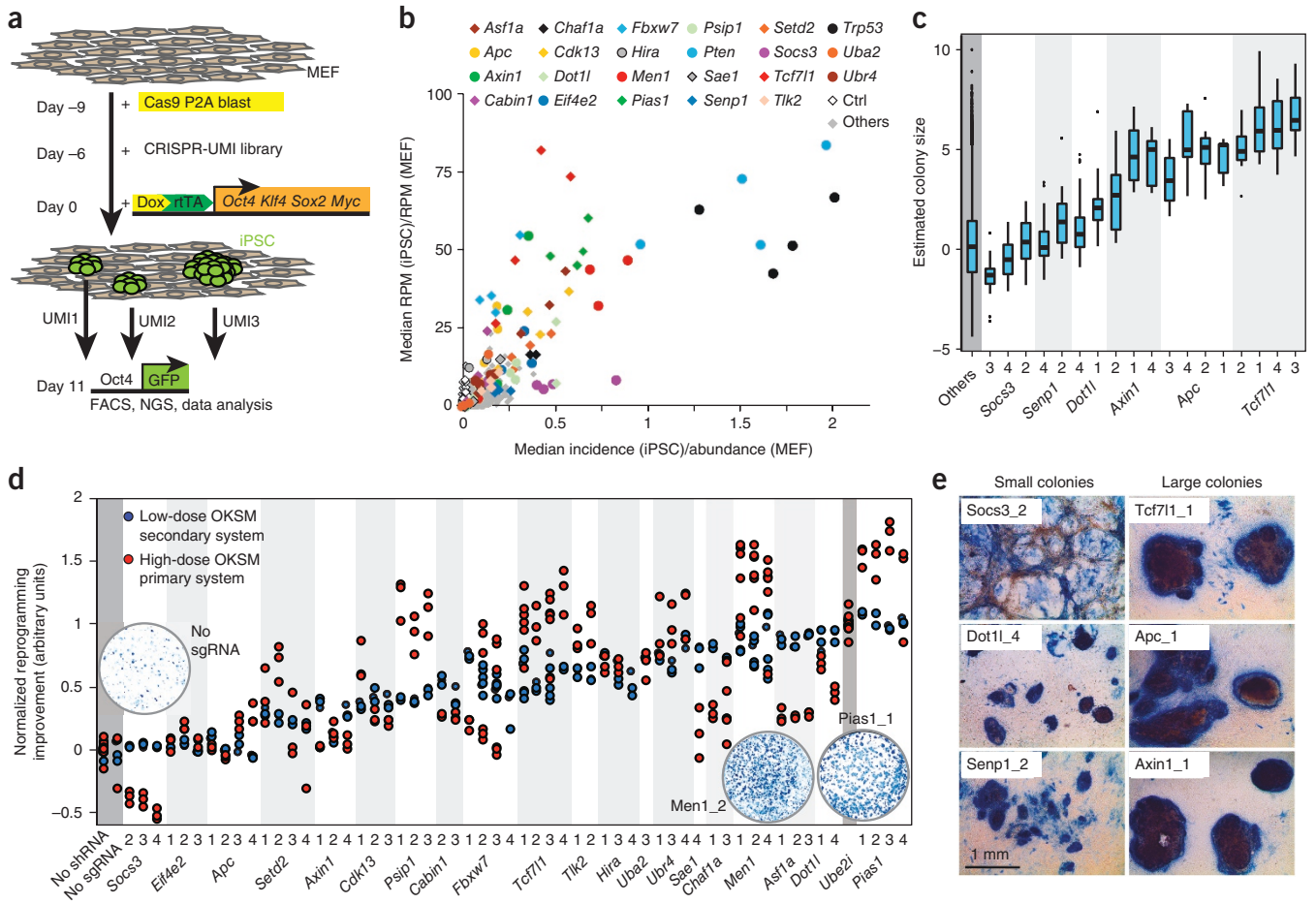


Figure 5 | CRISPR-UMI analysis of a positive-selection screen to identify roadblocks to reprogramming. (**a**) A schematic of the experimental setup. NGS, next-generation sequencing; rtTA, reverse tetracycline-controlled transactivator. (**b**) The enrichment of individual sgRNAs based on abundance (read counts; *y*-axis) and of individual sgRNAs based on incidence (colony number; *x*-axis). Both axes are normalized to the total number of reads (abundance) of the sgRNA in the untreated MEF population. The data points represent the median enrichment of sgRNAs in $n = 4$ experiments. Ctrl, control; RPM, reads per million mapped reads. (**c**) Colony size as estimated on the basis of the normalized and median scaled abundance of unique barcode–guide combinations. The number of colonies per guide is shown in **Supplementary Figure 10a**. Center line, median; hinges, 25th and 75th quartiles; whiskers, median ± 1.58× the interquartile range (IQR)/2. Individual data points represent outliers. (**d**) Validation of single sgRNAs based on Oct4–GFP reporter signal, normalized to control (0) and *Ube2i* knockdown (1). Knockdown of *Ube2i* resulted in >100-fold improvement of reprogramming in low-dose OKSM. Insets in the lower right corner show validaton of Men1 and Pias1 by alkaline phosphatase staining of iPSCs. (**e**) Representative colonies in the independent validation experiment on day 10 after Dox induction, stained for alkaline phosphatase activity, confirming colony-size predictions.

treatment for 7 d. We passaged and cultured cells without Dox to assay iPSC fate independent of exogenous OKSM, and then carried out FACS purification of Oct4–GFP$^+$ cells on day 11 to isolate iPSCs. Subsequently, we isolated genomic DNA from iPSCs and MEFs and determined sgRNA abundance (i.e., iPSC number) and the number of independent guide–barcode combinations (i.e., iPSC colonies). We identified many known roadblocks of reprogramming such as *Trp53* (ref. 22), *Pten*[23], *Dot1l*[24], *Socs3* (ref. 25), *Sae1*, *Uba2*, and *Chaf1a*[26] (**Fig. 5b**), as well as novel candidates. sgRNAs targeting *Senp1*, *Socs3*, and *Dot1l* scored primarily on the axis for incidents of iPSC formation, highlighting the additional resolution gained from this additional readout parameter. In contrast, colony size reflected the reprogramming speed and/or growth kinetics of resultant iPSC colonies. We therefore plotted the median iPSC colony size by quantifying the read count for each barcode-tagged individual colony, and observed a distribution that was gene specific and relatively reproducible between sgRNAs (**Fig. 5c**, **Supplementary Fig. 10a**).

We used two approaches for validation, as reprogramming is strongly dependent on the timing and dose of OKSM[26]: (1) low OKSM levels obtained with the Dox-inducible OKSM MEFs, and (2) high OKSM levels obtained by lentiviral OKSM delivery. Knockdown of *Ube2i* served as a positive control, which improved reprogramming >100-fold in this setting[26]. As expected, almost all sgRNAs enhanced reprogramming efficiency (**Fig. 5d**, **Supplementary Fig. 10b**). The effect size for each gene varied between the two reprogramming systems, in agreement with prior findings, suggesting system-specific roadblocks of the iPSC reprogramming process[26–28]. Targeting of genes such as *Pias1*, encoding an E3 SUMO-protein ligase, and *Men1*, encoding the transcriptional cofactor Menin, markedly outperformed all tested and previously identified roadblocks in a primary reprogramming regimen. To test whether the read count distribution measured

in the screen indeed reflected the number versus the size of iPSC colonies, we imaged colonies 10 d after Dox induction for guides that were predicted to generate particularly large or small colonies. Indeed, the observed colony size correlated with the expected size distribution from primary screen analysis (**Fig. 5e**, **Supplementary Fig. 10c**). Thus, single-cell-based CRISPR-UMI analysis in a positive selection can robustly identify hits and determine the number versus the effect size of single selection events predicting different biological mechanisms.

DISCUSSION

Genome-wide dropout screens, both RNA-interference-based and CRISPR-based, have advanced the functional annotation of the genome[29–31]. However, conventional analyses of CRISPR screens are limited to analyses of cell-population behavior during the screening process. This limitation was recently overcome through the linking of sgRNA identification within a cell to single-cell RNA-seq[32–35]. These methods represent an in-depth analysis of single cells, requiring thousands of sequencing reads per cell, which limits the total number of cells that can be analyzed. CRISPR-UMI closes a remaining gap with respect to conventional pooled CRISPR screens by measuring heterogeneity at the single-cell level, with one read identifying the sgRNA in each cell. Although CRISPR-UMI does not enable in-depth analysis of single-cell effects, it does enable the analysis of several orders of magnitude more cells, and thus genome-wide approaches rather than more focused assays.

Genetic and epigenetic variability can be inherent to a cell line or editing-dependent. CRISPR-UMI distinguishes individually labeled cells, thereby accounting for these variations and resulting in higher resolution of signal from noise. Importantly, it does not require a greater read depth or different deep-sequencing modalities. It also does not interfere with conventional analysis, and thus offers an additional and independent optional layer of insight to CRISPR screens. LOF mutations of essential genes by definition interfere with clonal outgrowth, and thus prevent the use of CRISPR-UMI for mapping of essential genes; however, subtraction of outlier clones remains an option for the mapping of essentialomes.

We applied CRISPR-UMI to a sensitizer screen for etoposide and identified all the expected genes in the NHEJ pathway, as well as unanticipated genes such as *Abcc1* and *Zfp451*, by both conventional and CRISPR-UMI analysis. *Zfp451* has recently been associated with cellular stress including DNA damage[36–38], but a direct role in DSB resistance has not been reported previously. Our finding suggests that chemical inhibition might generate therapeutically relevant synergy with etoposide. CRISPR-UMI uncovered *Rad9a* and *Erbb4*, which were previously associated with DNA damage response[15–20]. These were not identified in the conventional analysis, presumably because of wild-type clones that dominated the sequencing space. The elimination of outliers from conventional analysis also removed false positive hits, such as *Trim71* and *Ell*, from the list of top-scoring genes. Furthermore, median clone depletion more accurately predicted true biological effects compared with classical analysis, which suffers from a conceptual maximal level of measurable depletion.

CRISPR-UMI enabled scoring of the number of independent iPSC colonies formed in a single screen, and thus efficient identification of reprogramming roadblocks. CRISPR-UMI identified Pias1, an E3 ligase of SUMOylation, and Menin, neither of which was previously implicated in iPSC reprogramming. Interestingly, Menin has been associated with other lineage identity switches[39,40], which potentially suggests a general role for the protein in the maintenance of lineage identity. Future experiments are needed to test the relevance of these genes in the reprogramming of human cells.

By studying incidence and effect size (i.e., the number and size of independent iPSC colonies), we separated different biological functions directly from sequencing data obtained as a screen readout. Examples are sgRNAs targeting *Axin1*, *Apc*, and *Tcf7l1*, all implicated in Wnt signaling, that led to few but very big iPSC colonies. In contrast, targeting of expected hits *Dot1l* and *Socs3* (ref. 25), as well as targeting of *Senp1*, resulted primarily in an increased number of independent UMIs with few reads representative of many small iPSC colonies, as previously shown for *Dot1l*[24], and might thus have been missed in conventional analysis. We hypothesize that *Socs3* knockout boosts LIF-induced Stat3 signaling[41], which leads to increased reprogramming toward iPSCs followed by differentiation to trophoblast giant cells[42]. The same principle will be valuable in the study of other stochastic positive-selection events such as differentiation and metastasis. Moreover, this method can detect clonal effects in positive-selection screens and thereby avoid false positive calls due to, for example, double-infection of one cell with two guides. In conclusion, CRISPR-UMI even allows for a higher multiplicity of infection because it can assess the reproducibility of phenotypes on a clonal level. This allows for reduced experiment size or increased sample complexity.

CRISPR-UMI is independent of and complementary to ongoing efforts to optimize guide efficiency[7,43,44], Cas9 variants[45–49], and on-target specificity[44,50]. In fact, it will allow quantification of the efficiency of guides to generate LOF alleles and thereby fuel the optimization of the 'screening hardware'. This increased resolution will be vital in more challenging screens, for example, with less robust negative-selection pressure. Moreover, we predict that single-cell-based analysis of screening outcomes will be critical in scenarios where clonal variance in the screening population might be high, such as in primary cells, differentiation, organoid culture, or, most noteworthy, *in vivo*. With rapidly increasing deep-sequencing power, this method will gain even further depth and resolution.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We acknowledge everybody involved in the generation of data and of the manuscript. We thank J. Zuber (Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Vienna, Austria) for the retroviral backbone used to generate the guide library, and K. Hochedlinger (Department of Molecular Biology, Cancer Center and Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; Department of Stem Cell and Regenerative Biology and Harvard Stem Cell Institute, Cambridge, Massachusetts, USA; Howard Hughes Medical Institute, Chevy Chase, Maryland, USA) for provision of the Dox-inducible reprogramming system. We thank A. Stark for important suggestions and critical reading of the manuscript, J. Jude for critical

discussion and sharing of protocols, and P. Svoboda for discussion and advice. We are grateful to all IMBA/IMP services, in particular bioinformatics, biooptics, molecular biology, media kitchen, and graphics for technical support, as well as to J. Brennecke and A. Andersen (Life Science Editors) for critical reading and editing of the manuscript. We thank the VBCF NGS facility. This work was supported by IMBA, the Austrian Academy of Sciences (OEAW), Novartis Institute of Biomedical Research, and AstraZeneca.

AUTHOR CONTRIBUTIONS

G.M. and U.E. conceived the study. G.M. cloned the library and performed the etoposide screens and bioinformatic studies. M.H. generated the Cas9 inducible ESC line and supported follow-up experiments. E.B. and U.E. performed the iPSC screen, and S.-H.W., G.V., and U.E. validated it. T.R.B. and M.N. performed bioinformatic analyses. S.Z., Y.L., and D.S. supported experiments. M.A. generated the vector backbone for the sgRNA library. J.R.-H., R.N., and D.H. supported the design and generation of sgRNA libraries. U.E. wrote the manuscript with support from G.M., D.S., D.H., and all other co-authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Koike-Yusa, H., Li, Y., Tan, E.-P., Del Castillo Velasco-Herrera, M. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
2. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
3. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
4. Hart, T. *et al.* High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
5. Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
6. Graham, D.B. & Root, D.E. Resources for the design of CRISPR gene editing experiments. *Genome Biol.* **16**, 260 (2015).
7. Doench, J.G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
8. Shalem, O., Sanjana, N.E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
9. Miles, L.A., Garippa, R.J. & Poirier, J.T. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens. *FEBS J.* **283**, 3170–3180 (2016).
10. Burden, D.A. *et al.* Topoisomerase II etoposide interactions direct the formation of drug-induced enzyme-DNA cleavage complexes. *J. Biol. Chem.* **271**, 29238–29244 (1996).
11. Jackson, S.P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
12. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
13. Black, S.J., Kashkina, E., Kent, T. & Pomerantz, R.T. DNA polymerase θ: a unique multifunctional end-joining machine. *Genes (Basel)* **7**, 67 (2016).
14. Takata, K., Reh, S., Tomida, J., Person, M.D. & Wood, R.D. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat. Commun.* **4**, 2338 (2013).
15. Gilmore-Hebert, M., Ramabhadran, R. & Stern, D.F. Interactions of ErbB4 and Kap1 connect the growth factor and DNA damage response pathways. *Mol. Cancer Res.* **8**, 1388–1398 (2010).
16. Icli, B., Bharti, A., Pentassuglia, L., Peng, X. & Sawyer, D.B. ErbB4 localization to cardiac myocyte nuclei, and its role in myocyte DNA damage response. *Biochem. Biophys. Res. Commun.* **418**, 116–121 (2012).
17. Mukherjee, B., Choy, H., Nirodi, C. & Burma, S. Targeting nonhomologous end-joining through epidermal growth factor receptor inhibition: rationale and strategies for radiosensitization. *Semin. Radiat. Oncol.* **20**, 250–257 (2010).
18. Greer Card, D.A., Sierant, M.L. & Davey, S. Rad9A is required for G2 decatenation checkpoint and to prevent endoreduplication in response to topoisomerase II inhibition. *J. Biol. Chem.* **285**, 15653–15661 (2010).
19. He, W. *et al.* A role for the arginine methylation of Rad9 in checkpoint control and cellular sensitivity to DNA damage. *Nucleic Acids Res.* **39**, 4719–4727 (2011).
20. Smilenov, L.B. *et al.* Combined haploinsufficiency for ATM and RAD9 as a factor in cell transformation, apoptosis, and DNA lesion repair dynamics. *Cancer Res.* **65**, 933–938 (2005).
21. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat. Methods* **7**, 53–55 (2010).
22. Marión, R.M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460**, 1149–1153 (2009).
23. Liao, J. *et al.* Inhibition of PTEN tumor suppressor promotes the generation of induced pluripotent stem cells. *Mol. Ther.* **21**, 1242–1250 (2013).
24. Onder, T.T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–602 (2012).
25. Buckley, S.M. *et al.* Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell* **11**, 783–798 (2012).
26. Cheloufi, S. *et al.* The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).
27. dos Santos, R.L. *et al.* MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. *Cell Stem Cell* **15**, 102–110 (2014).
28. Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65–70 (2013).
29. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
30. Rauscher, B., Heigwer, F., Breinig, M., Winter, J. & Boutros, M. GenomeCRISPR—a database for high-throughput CRISPR/Cas9 screens. *Nucleic Acids Res.* **45**, D679–D686 (2017).
31. Wang, T. *et al.* Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* **168**, 890–903 (2017).
32. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
33. Dixit, A. *et al.* Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
34. Adamson, B. *et al.* A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
35. Jaitin, D.A. *et al.* Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896 (2016).
36. Koidl, S. *et al.* The SUMO2/3 specific E3 ligase ZNF451-1 regulates PML stability. *Int. J. Biochem. Cell Biol.* **79**, 478–487 (2016).
37. Guzzo, C.M. *et al.* RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage. *Sci. Signal.* **5**, ra88 (2012).
38. Cappadocia, L., Pichler, A. & Lima, C.D. Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase. *Nat. Struct. Mol. Biol.* **22**, 968–975 (2015).
39. Lu, J. *et al.* Alpha cell-specific Men1 ablation triggers the transdifferentiation of glucagon-expressing cells and insulinoma development. *Gastroenterology* **138**, 1954–1965 (2010).
40. Liu, L. *et al.* Targeting Mll1 H3K4 methyltransferase activity to guide cardiac lineage specific reprogramming of fibroblasts. *Cell Discov.* **2**, 16036 (2016).
41. Boyle, K. *et al.* Deletion of the SOCS box of suppressor of cytokine signaling 3 (SOCS3) in embryonic stem cells reveals SOCS box-dependent regulation of JAK but not STAT phosphorylation. *Cell. Signal.* **21**, 394–404 (2009).
42. Takahashi, Y. *et al.* SOCS3: an essential regulator of LIF receptor signaling in trophoblast giant cell differentiation. *EMBO J.* **22**, 372–384 (2003).
43. Heigwer, F. *et al.* CRISPR Library Designer (CLD): software for multispecies design of single guide RNA libraries. *Genome Biol.* **17**, 55 (2016).
44. Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
45. Richter, F. *et al.* Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res.* **44**, 10003–10014 (2016).
46. Kleinstiver, B.P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
47. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
48. Larson, M.H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **8**, 2180–2196 (2013).
49. Hess, G.T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
50. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).

ONLINE METHODS

A detailed step-by-step protocol for the methods described in this paper is available at *Protocol Exchange*[51].

Guide selection. sgRNAs targeting mouse nuclear genes as well as drugged orthologs and a set of hand-selected genes with four sgRNAs per gene (five sgRNAs per gene for the subset of drugged genes) were selected via a bioinformatics pipeline. We aimed to design a guide-selection algorithm that would take both guide efficiency and biological effect due to gene structure into account. The basis of the guide selection was the activity score as described by Doench *et al.*[7]. Additionally, we identified properties of each guide and exon under consideration and penalized the Doench score accordingly. We identified all exonic PAM sites in mouse genome mm10 (ref. 52). We excluded sgRNAs that were incompatible with our cloning strategy (i.e., those that contained GAAGAC, GTCTCC, CTCGAG, CGTCTC, or GAGACG; started with AAGAC; or ended with CTCGA). We then calculated Doench scores for all potential sgRNAs. We penalized the Doench scores according to heuristic rules (exact penalty scores can be found in **Supplementary Fig. 2**) with the aim of selecting the sgRNAs most likely to lead to LOF phenotypes. Those rules included exon properties such as the presence or absence of protein domains annotated in the Pfam database[53], exon size, and whether or not exon length is a multiple of 3 bp. Then we created penalties for exon distribution, in order to spread sgRNAs over many exons, where only the sgRNA with the best Doench score per exon did not get penalized. We also avoided sgRNAs that were less than 4 nt away from another, better scoring sgRNA. Furthermore, we penalized sgRNAs that cut DNA upstream of a possible alternative ATG start codon and sgRNAs that cut in exons that were not common to all annotated transcripts from that locus. We avoided sgRNAs that contained a stretch of four or more threonines in a row, which would act as Pol-III terminators. We calculated a distance penalty based on the distance from the sgRNA to the transcriptional start, ranging from 1 to 0.5. Then we calculated a simple off-target prediction (**Supplementary Fig. 2**) against all exonic sequences containing a PAM site. The off-target prediction scores weighted mismatches by position in the sgRNA sequence[54,55]. We re-ranked the penalized Doench scores including the off-target analysis and picked the top four sgRNAs per gene (the top five sgRNAs for druggable genes) for chip oligo synthesis (CustomArray Inc.). For negative control guides we used a published list of human control guides[3] and removed all guides that had a perfect match against the mouse genome. We included a total of 112 control guides in our mouse library, targeting 6,560 genes (**Supplementary Table1**).

Library cloning. We ordered a gBlock (IDT) flanked by primer binding sites for amplification, and restriction sites *Eco*RI and *Mfe*I for cloning of the Illumina i7 primer binding site followed by a 10-bp random nucleotide sequence and the Illumina P7 adaptor (acgatgagcagag ccagaaccagaaggaacttgactctagaGATCGGAAGAGCACACGT CTGAACTCCAGTCACNNNNNNNNNNgtcctcatctgagag ctactcatcaacggtATCTCGTATGCCGTCTTaTGCTTGTTA ATTAAGAATTCctggacga) (note: we changed C to A in the P7 adaptor sequence to eliminate a *Bbs*I restriction site in the adaptor for library cloning, but reintroduced the C during PCR in the DNA-sample prep before next-generation sequencing (NGS)). The gBock was digested with *Eco*RI (NEB; R3101L) and *Mfe*I (NEB; R3589L), purified on a column (Qiagen; 27106), and precipitated with ethanol. Vector backbone (**Supplementary Fig. 4a**) was digested with *Xba*I (NEB; R0145L) and *Mfe*I (NEB; R3589L) and dephosphorylated with rSAP (NEB; M0371L); a 1.5-kb stuffer containing an *Eco*RI restriction site was excised, and vector backbone fragments were separated by agarose gel electrophoresis, gel extracted (Qiagen; 28704), and precipitated with ethanol. 2 µg of vector were ligated with 125 ng of insert at a vector:insert molar ratio of 1:3, with 2 µl of T4 DNA ligase (NEB; M0202M) in a total volume of 200 µl split into ten reactions of 20 µl each. The ligation was purified on a column (Qiagen; 27106) and electroporated into electrocompetent XL-1 Blue cells (Agilent; 200249), 80 µl in a 0.2-cm cuvette at 2.5 kV (Bio-Rad; Gene Pulser II); on the basis of colony count the electroporation complexity was estimated as 1 million. In a second library-cloning step, sgRNAs were cloned into vector containing complex barcodes. The vector was digested with *Bbs*I (NEB (R0539L) or Fermentas (ER1011)). A stuffer containing an *Xho*I binding site was excised and dephosphorylated with rSAP (NEB; M0371L), and linear fragments were isolated by agarose gel electrophoresis and ethanol-precipitated. sgRNAs were ordered on a chip (CustomArray Inc.), and subsets of the oligos were amplified with specific flanking primers with ten cycles of PCR. PCR product was purified on a column (Qiagen; 27106) and digested overnight with *Bbs*I (NEB (R0539L) or Fermentas (ER1011)). Vector and insert were ligated in a Golden Gate reaction with 0.25 µl of T4 DNA ligase (NEB; M0202M) and 1 µl of *Bbs*I (NEB (R0539L) or Fermentas (ER1011)) in a 50-µl reaction volume. The reaction was cycled 20 times (one cycle consisted of 5 min at 37 °C and 5 min at 16 °C) and then subjected to a 10-min inactivation at 50 °C. Plasmids were purified on columns (Qiagen; 27106), ethanol-precipitated, and electroporated into electrocompetent XL-1 Blue cells (Agilent; 200249). Electroporated XL-1 Blue cells were collected in 2 ml of recovery diluent, incubated at 37 °C and 200 r.p.m. for 40 min, plated on two square 245 × 245 mm LB agar plates containing 100 µg/ml ampicillin (LB–amp; Thermo; 166508), and incubated at 37 °C for 10 h. Bacteria were collected from the plates and grown in 2 L of LB–amp at 100 µg/ml (Sigma; A9518) for 2–4 h until the OD reached 2.0. Plasmid DNA was prepared (Macherey-Nagel NucleoBond Xtra Maxi Kit) according to the manufacturer's recommendations.

Embryonic stem cell culture. We used an mESC clone derived from a derivative of HMSc2, called AN3-12, with Dox-inducible Cas9 (T3G-Cas9-IRES-mCherry PGK-GFP-rtTA) for this study. The following ESC medium (ESCM) was used: 450 ml of DMEM (Sigma; D1152), 75 ml of FCS (Gibco), 5.5 ml of penicillin–streptomycin (P/S; Sigma; P0781), 5.5 ml of NEAA (Sigma; M7145), 5.5 ml of L-glutamine (Sigma; G7513), 5.5 ml of sodium pyruvate (Sigma; S8636), 0.55 ml of β-mercaptoethanol (Merck, 805740; 10 µl diluted in 2.85 ml of PBS for a 1,000× stock), 7.5 µl of LIF (2 mg/ml). Cell-culture-grade dishes were from Greiner (15 cm; 639160) and NUNC (all other formats—for example, 10-cm dish, Nunclon Surface, cat no. 150350; six-well Nunclon Surface, cat no. 140675). Cells were trypsinized and replated every second day and frozen in FCS:ESCM:DMSO at 4.5:4.5:1 (vol/vol). Cells were tested for mycoplasma every second week. For etoposide

doi:10.1038/nmeth.4466

treatment, medium was supplemented every day with 3.3 nM etoposide, an LD_{30} dose (i.e., the dose expected to kill 30% of treated cells), for 8 d of treatment (Sigma; E2600000); 1,000× etoposide stocks dissolved in PBS–10% EtOH were used. For Dox treatment, medium was supplemented every day with 1 µg/ml Dox (Sigma). Cells were tested for mycoplasma contamination every second week.

Viral vectors and embryonic stem cell infection. For retroviral library generation, barcoded CRISPR library virus carrying a neomycin-resistance cassette was packaged in PlatinumE cells (Cell Biolabs) according to the manufacturer's recommendations. Virus-containing supernatant was frozen at −80 °C. 300 million ESCs were infected with a 1:10 dilution of virus-containing supernatant for 24 h in the presence of 2 µg/ml polybrene (Sigma; TR-1003). 24 h after infection, selection for infected cells was started with G418 (Gibco) at 0.5 mg/ml. To estimate the multiplicity of infection, we plated 10,000 cells on 15-cm dishes and selected with G418. For comparison, we plated an additional 1,000 cells and did not expose them to G418 selection. On day 10, colonies were counted. After 24 h of selection, cells were split and 480 million cells were seeded on 60 15-cm dishes (Greiner; 639160). After that, cells were kept at a minimum number of 300 million cells during editing and screening.

iPSC screen and validation. MEFs containing *Col1a1::tetOP-OKSM*, *Oct4-Gfp*, and *Rosa26 M2rtTA* alleles or Oct4–GFP alone were harvested from embryos at embryonic day 13.5 (E13.5)[21]. iPSCs were derived in DMEM supplemented with 15% FBS, 100 U/ml penicillin, 100 µg/mlstreptomycin, sodium pyruvate (1 mM), L-glutamine (4 mM), 1,000 U/ml LIF, 0.1mM β-mercaptoethanol, and 50 µg/ml ascorbic acid at 37 °C, 5% CO_2, 4.5% O_2. MEFs were infected with a lentiviral vector delivering Cas9, selected for blasticidin resistance for 3 d, and subsequently infected with the sgRNA library. MEFs were treated with 0.5 mg/ml G418 for 3 d and 0.25 mg/ml G418 for an additional 3 d. For iPSC induction, MEFs were plated at densities of 500,000 cells per 15-cm dish and induced with doxycycline (1 µg/ml) for 7 d. After passaging for an additional 4 d in doxycycline-free media, Oct4–GFP-expressing cells were sorted from each replicate in a FACSAria III (BD Bioscience). All validation experiments were performed in six-well dishes in triplicate, starting with 20,000 MEFs (Dox induction) or 40,000 cells (OKSM infection). Primary reprogramming was done by infection with a lentiviral vector carrying OKSM factors as well as puromycin resistance, followed by selection for puromycin resistance for 3 d. Oct4 expression was quantified with a FACS BD LSR Fortessa (BD Biosciences), and data were analyzed with FlowJo. Cells were tested for mycoplasma contamination every second week.

Etoposide-hypersensitivity validations. For hit validation we used mESCs expressing Cas9 under the control of doxycycline (T3G-Cas9-IRES-mCherry PGK-GFP-rtTA). We generated four knockout clones for the genes *Lig4*, *Zfp451*, *Slc25a4*, *Adcy3*, *Rad9a*, and *Erbb4*, and two knockout clones for the genes *Trim71* and *Rac1* (sgRNAs and PCR primers for genotyping are in **Supplementary Table 2**), and we tested every clone in triplicate. We pretreated cells for 7 d with 5 nM etoposide (Sigma; E2600000) and then measured the change in cell viability (treated versus control) within a 3-d selection with 10 nM etoposide. Viability was measured by Alamar Blue staining (DAL1100; Thermo Fisher) according to the manufacturer's recommendations.

DNA harvest and NGS sample preparation. Readout of pooled CRISPR screens relies on precise PCR amplification of the integrated sgRNA cassette. However, in a genomic DNA prep, the sgRNA cassette makes up only about 0.1 p.p.m. of the total DNA. We improved this ratio by three to four orders of magnitude by flanking our sgRNA cassette with *Pac*I sites (**Fig. 2b**) and carrying out a size-selective precipitation with digested genomic DNA. After 8 d of selection we lysed 170 million cells per condition in SDS lysis buffer (10 mM Tris, pH 8, 1% SDS, 10 mM EDTA, 100 mM NaCl) plus 1 mg/ml Proteinase K (Sigma; P4032) and RNaseA (Qiagen). Genomic DNA was purified by phenol extraction and precipitated with isopropanol. Samples were digested with *Pac*I (NEB; R0547L). Size-selective precipitation was carried out with Speed Beads (GE45152105050250; Sigma-Aldrich) according to the manufacturer's recommendations. Each Sample was PCR-amplified in 200 individual 50-µl PCR reactions with primers (AATGATACGGCGACCACCGA GATCTACAC-NNNNNN-CGAGGGCCTATTTCCCATGATT CCTTC, where the 6-bp NNNNNN sequence represents specific experimental indices used for demultiplexing samples after NGS, and CAAGCAGAAGACGGCATACGAGATACCGTTGA TGAGTAG). For qPCR analysis, we used GoTaq qPCR master mix (A6001; Promega) with primers AATGATACGGCGACC ACCGAGATCTACACGAGTGGCGAGGGCCTATTTCCCA TGATTCCTTC and CAAGCAGAAGACGGCATACGAGATA CCGTTGATGAGTAG for detection of a 579-bp amplicon on the 589-bp *Pac*I fragment with the CRISPR-UMI cassette, and GCCTTTAAGCCAATGCTAGCTG and GTAAATGGACAGA GGGTGTTTAACC as a control with a 582-bp amplicon on a 7,710-bp *Pac*I fragment. PCR samples were purified on columns (Qiagen; 27106) and pooled and size-separated by agarose gel electrophoresis. The 589-bp band was excised and purified on a mini-elute column (Qiagen; 28204). This sample was sequenced on an Illumina HiSeq 2500 in a single-read 50-dual-indexing sequencing run. We sequenced sgRNA sequences with a tenfold-concentrated custom read primer (CGATTTCTTGGCTTTATA TATCTTGTGGAAAGGACGAAACACCG). For analysis, it is necessary to obtain at least 10 bp for index 1 (barcode) and 6 bp for index 2 (experimental index).

Data analysis. For data analysis, we assigned sgRNA, UMI sequences, and experimental indices to all reads using Bowtie, SAMtools, FASTX-Toolkit, and custom scripts. All scripts and details are available online at http://mendel.imp.ac.at/SEQUENCES/Michlits/ and will be published in *Protocol Exchange*.

For the representation of guides in the library (**Fig. 2a**), library subpools were sequenced with individual experimental indices, and the reads within each subpool were normalized to the median guide read of all subpools.

For conventional CRISPR data analysis (**Fig. 3a–c**, **Supplementary Fig. 5d**, we calculated the depletion or enrichment of guides or clones using equation (1) and added a pseudocount of 0.5 if there were no reads for a guide. *P* values were calculated using cumulative binominal distribution functions (scipy.stats 0.19.0) as in equation (2). We plotted depletion (*x*-axis) against p_{cdf} (*y*-axis) in a volcano plot.

$$d = \frac{\mathrm{RPM}_{\mathrm{treated}}}{\mathrm{RPM}_{\mathrm{control}}} \quad (1)$$

where d is the depletion of the guide or clone, and RPM is the reads per million mapped reads of that guide or clone.

$$p_{\mathrm{cdf}} = \sum_{i=0}^{x} \binom{n}{x} p^x (1-p)^{n-x} \tag{2}$$

where p_{cdf} is the P value for the cumulative binominal distribution function, x is the number of reads of the guide or clone in the experiment, n is the total number of reads in the experiment, and p is the probability (reads of guide or clone in the control versus the total reads in the control).

For clonal CRISPR-UMI analysis (**Fig. 3d**, **Supplementary Fig. 5e**) we generated a 'volcano-like' plot. We considered clones that contained at least three total reads (in treated and control samples combined) and guides with at least five clones present. To account for random barcode sequencing errors in experiments, we used adjacency-based barcode collapse, given an edit distance of 1 and a threefold read-count difference, using custom scripts. To combine data on the clone level with data on the guide level, we plotted the median clone depletion against the MAGeCK 'neg score', a score for depletion for an sgRNA computed with MAGeCK 0.5.5 (ref. 12). Instead of using MAGeCK to calculate the depletion score of a gene on the basis of n guides, we used it to calculate depletion scores of a guide on the basis of n clones. Plotting the median depletion (x-axis) against MAGeCK neg scores (y-axis) produced a volcano-like plot.

To compare the performance of conventional analysis to that of CRISPR-UMI analysis on the guide level (**Fig. 4a**, **Supplementary Fig. 8a**), we ranked guides by both depletion and P values for conventional analysis and by median depletion and MAGeCK neg scores for CRISPR-UMI analysis, and then generated a combined guide score using equation (3).

$$\mathrm{GS} = \frac{\mathrm{Rank}_{\mathrm{depletion}}}{N_{\mathrm{total\ guides}}} \times \frac{\mathrm{Rank}_{p\ \mathrm{value}}}{N_{\mathrm{total\ guides}}} \tag{3}$$

where GS is the guide score, a combined score for the evaluation of guides; $\mathrm{rank}_{\mathrm{depletion}}$ is the rank of the guide by depletion (conventional analysis) or median depletion (CRISPR-UMI analysis); $\mathrm{rank}_{P\ \mathrm{value}}$ is the rank of the guide by P value (conventional analysis) or MAGeCK neg score (CRISPR-UMI analysis); and $N_{\mathrm{total\ guides}}$ is the total number of guides in the analysis.

For gene ranking (**Fig. 3d**) we used MAGeCK for conventional analysis and combined guide scores for CRISPR-UMI analysis with Fisher's method of P value combination. For the top five genes with the lowest P values we could not apply Fisher's method because of numerical restrictions, so we sorted those five genes by combining P values using equation (4).

$$P_{\mathrm{gene}} = \sqrt[n]{\prod_{i=1}^{n} \mathrm{GS}_i} \tag{4}$$

where P_{gene} is the score for the gene, GS_i is the guide score of guide i, and n is the total number of guides for that gene.

To calculate SNRs for a screen (**Fig. 3c**,d, **Supplementary Fig. 5e**) we defined the signal of a guide as the distance from the origin of a volcano plot. This considers separation from noise in both depletion (x-axis) and significance (y-axis). In volcano plots the x-axis indicated the ratio of guide reads in treated samples to those in the control (for CRISPR-UMI, the median of many clones), and the y-axis indicated the negative logarithm of the P value as determined by binominal distribution in conventional analysis or the negative score as determined by MAGeCK for multiple clones per guide in CRISPR-UMI analysis. Distance on the x-axis was normalized to the guide with the strongest depletion in the experiment. Distance on the y-axis was normalized to the guide with best significance score. Signal was the diagonal distance of the x-normalized and y-normalized distance from the origin calculated with equation (5). SNRs were calculated with equation (6).

$$S_i = \sqrt{\left(\frac{1-d_i}{1-d_{\mathrm{min}}}\right)^2 + \left(\frac{\log_{10}(P_i)}{\log_{10}(P_{\mathrm{min}})}\right)^2} \tag{5}$$

where S_i is the signal of guide i, d_i is the depletion of guide i (conventional) or the median depletion of clones for guide i (CRISPR-UMI), d_{min} is the depletion of the most strongly depleted guide in the comparison (conventional) or the lowest median depletion of all guides in the comparison (CRISPR-UMI), P_i is the P value for depletion for guide i (conventional) or the MAGeCK neg score for guide i (CRISPR-UMI), and P_{min} is the lowest P value for all guides in the comparison (conventional) or the lowest MAGeCK neg score for all guides in the comparison (CRISPR-UMI).

$$\mathrm{SNR} = \frac{\bar{S}_{\mathrm{NHEJ}}}{\sigma_{\mathrm{CTRL}}} \tag{6}$$

where SNR is the signal-to-noise ratio, $\bar{S}_{\mathrm{NHEJ}}$ is the average signal of all guides of the NHEJ pathway (*Lig4*, *Nhej1*, *Xrcc4–6*), and σ_{CTRL} is the s.d. of the signal from all control guides.

We evaluated CRISPR-UMI in comparison with the conventional screen by using a value termed 'depletion(NHEJ/control)' (**Fig. 3c**,d). This value is a ratio of the geometric mean of the fold change of all guides versus NHEJ-pathway genes (*Lig4*, *Nhej1*, *Xrcc4–6*) and the geometric mean of the fold change of nontargeting control guides.

To assess the efficiency of different screening and analysis methods (**Fig. 4c**), we evaluated the ranking of guides by guide scores calculated from equation (3). We determined the average number of guides present per gene among the top guides (y-axis) while expanding the list of top guides from 1 to 100 (x-axis). For example, if among the top 30 guides (x-axis) there are guides against 15 different genes, there are on average 2 (y-axis) guides per gene.

For incidence versus abundance analysis of a positive-selection screen of MEF reprogramming to iPSCs (**Fig. 5b**), we determined read counts for all individual UMIs. Read counts for all UMIs belonging to the same guide were added up to provide a measure of abundance for each guide. The number of independent UMIs per guide that passed the threshold for an iPSC colony (at least ten reads) was a measure of incidence. Enrichment for guides in either abundance or incidence was normalized to the abundance of the guides in the starting MEF population. Each data point shown in **Figure 5b** is the median of four separate experiments carried out with MEFs derived from four embryos.

To estimate iPSC colony size (**Fig. 5c**, **Supplementary Fig. 10a**), we determined read counts for all UMIs and scaled them by library size across eight treated and four untreated replicates. UMI barcodes with a hamming distance of 1 were collapsed to account for UMI sequencing errors using UMI-tools[56]. We filtered UMIs with low counts by retaining UMIs with five or more read

doi:10.1038/nmeth.4466

counts. Starting with the highest represented guide–UMI combination, we included all UMIs up to a cumulative percentage of 90% per guide. Normalized UMI counts were divided by the median of the UMI counts for each sample, and the values from the eight treated replicates were pooled, log-transformed, and used to visualize the distribution of the abundance estimates relative to the experimental median for each guide.

A detailed description of the experimental design, software, materials, and reagents is available in the **Life Sciences Reporting Summary**.

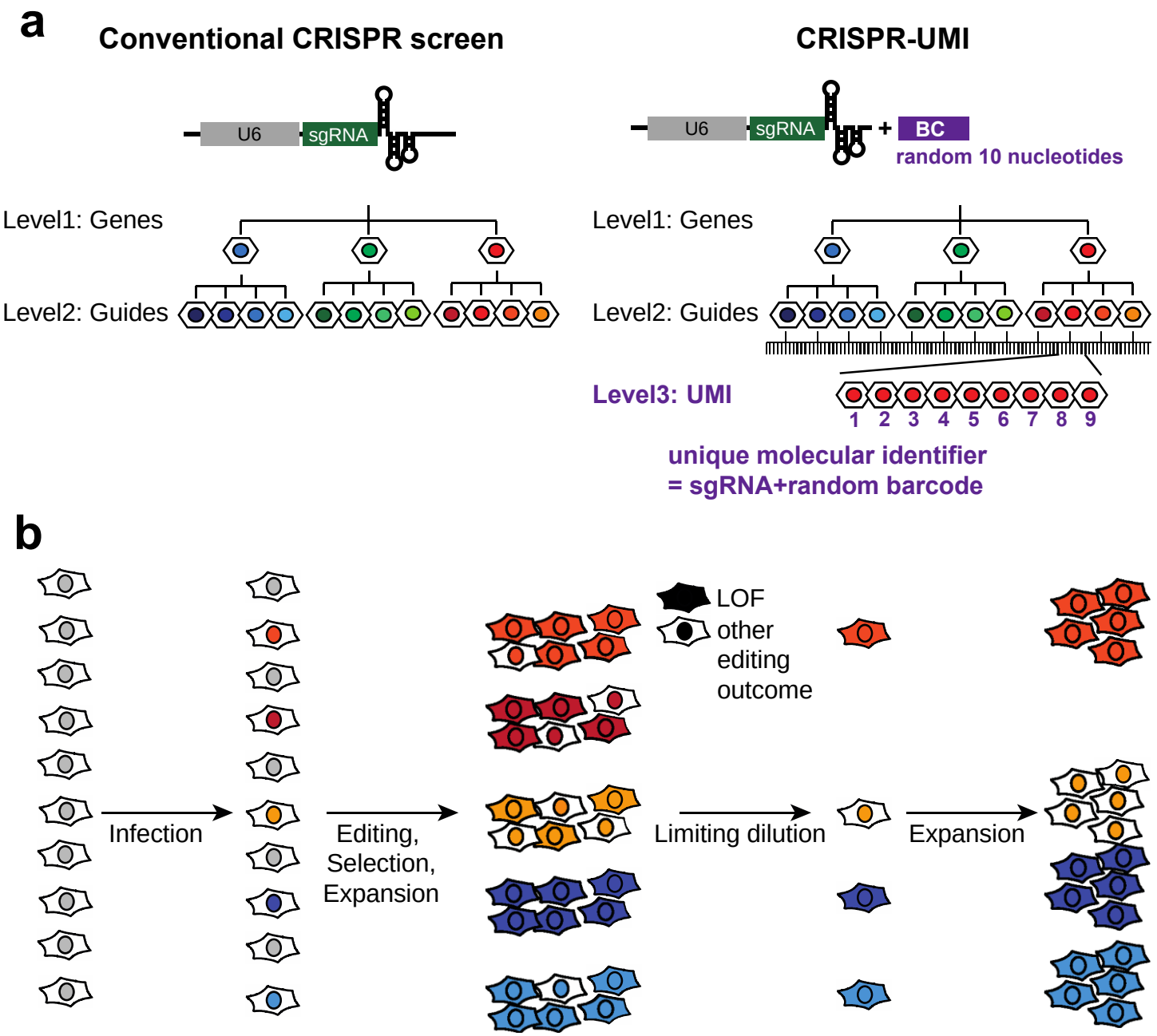
Statistical information. All error bars in figures represent s.d. or s.e.m. unless otherwise stated. Statistical information for conventional and CRISPR-UMI screen analysis is described above in "Data analysis."

Code availability. All codes and scripts used for guide selection and data analysis are available online at http://mendel.imp.ac.at/SEQUENCES/Michlits/ and have been uploaded to *Protocol Exchange*[51].

Step-by-step protocol. A detailed step-by-step protocol including all necessary steps of the CRISPR-UMI screens is available online at http://mendel.imp.ac.at/SEQUENCES/Michlits/ and was uploaded to *Protocol Exchange*[51].

Data availability. All genomic data are available via the NCBI Sequence Read Archive under accession code PRJNA383356. Experimental indices for NGS data are available in **Supplementary Table 3**. Source data files for **Figures 2**–**5** and **Supplementary Figures 4**, **5** and **7**–**10** are available online.

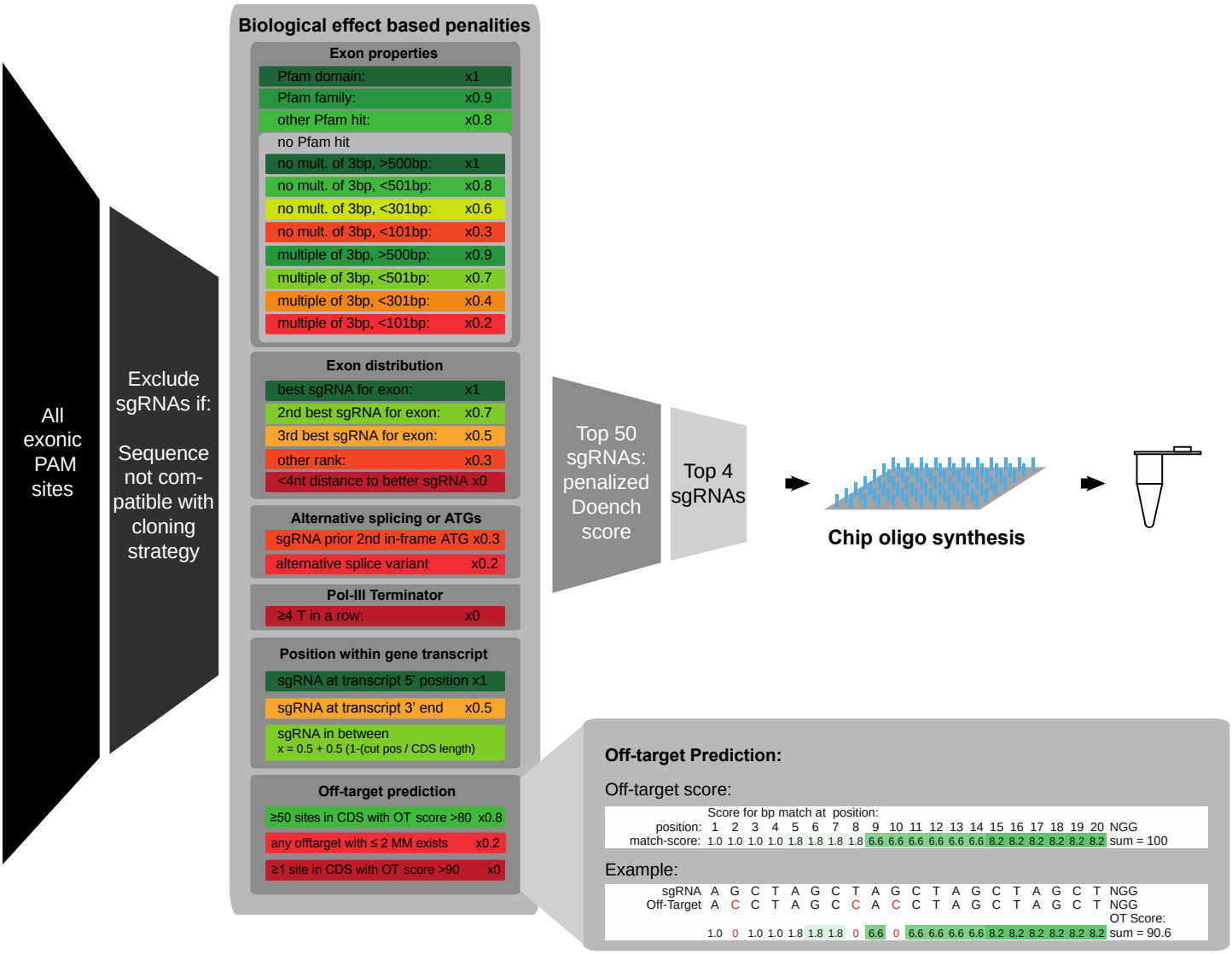
51. Michlits, G., Burkard, T.R., Novatchkova, M. & Elling, U. CRISPR-UMI step-by-step: a protocol for robust CRISPR-screening. http://dx.doi.org/10.1038/protex.2017.111 (2017).
52. Rosenbloom, K.R. *et al.* The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).
53. Finn, R.D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
54. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).
55. Tsai, S.Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
56. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).



Supplementary Figure 1

A schematic illustration of UMI use in CRISPR–Cas9 screens.

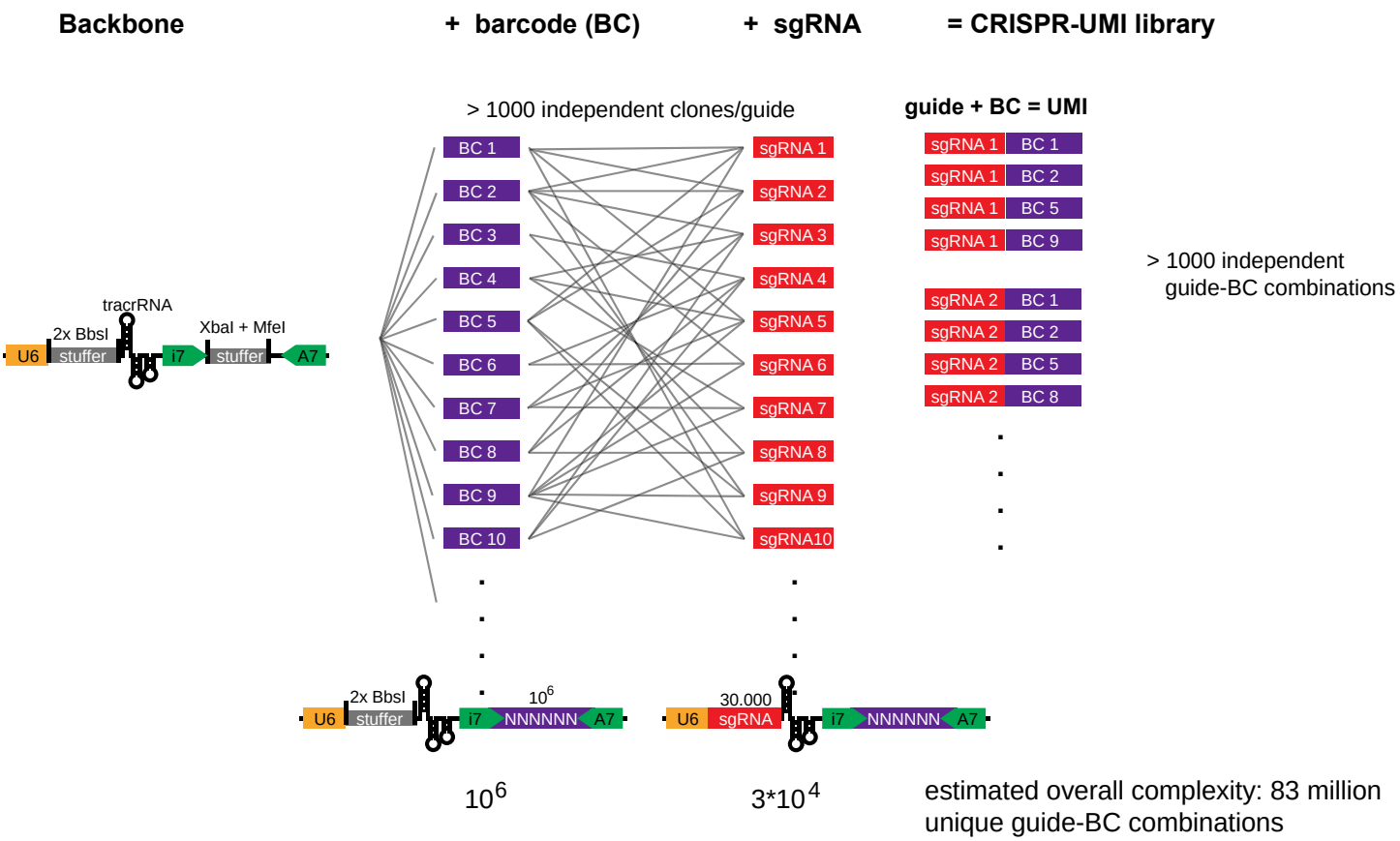
(a) Data analysis in CRISPR screens is conventionally based on several sgRNAs targeting the same gene. Introduction of random barcodes at complexities well above total analyzed cell number will tag each individual cell with a unique molecular identifier (UMI). This generates a third layer of information at single cell level, generating biological replicas. (b) Schematic illustration of the generation of single cell derived clones using a limiting dilution followed by clonal expansion.



Supplementary Figure 2

The bioinformatic pipeline of sgRNA prediction.

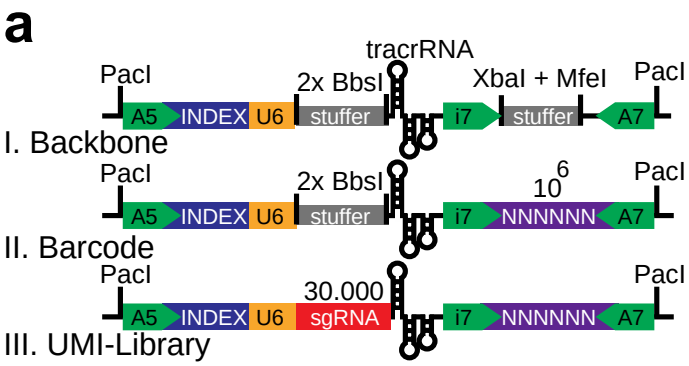
Doench-scores as measure of predicted sgRNA activity were calculated for all exonic sgRNAs compatible with our cloning strategy. Doench scores were penalized based on a ruleset for biological effects. Those rules combine evaluation of exon length, prediction of protein domains, alternative splicing and ATG start codons, Pol-III terminator sequences, position of the sgRNA within the CDS. The penalties also spread selected sgRNAs over different exons and include off-target prediction penalties.



Supplementary Figure 3

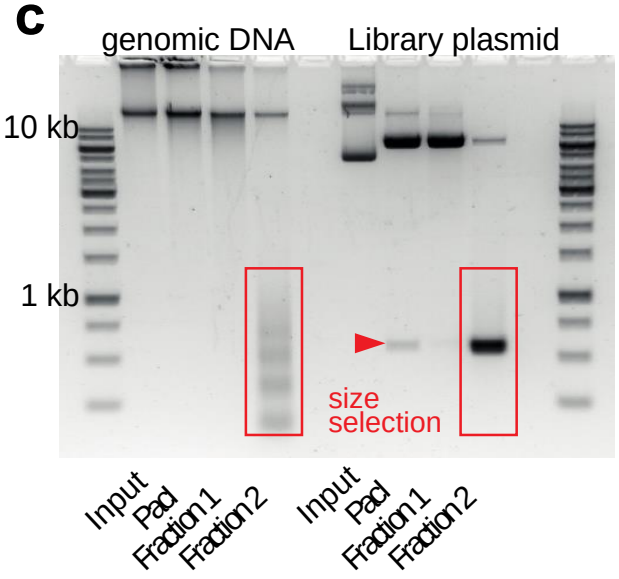
A scheme illustrating the generation of CRISPR-UMI library complexity.

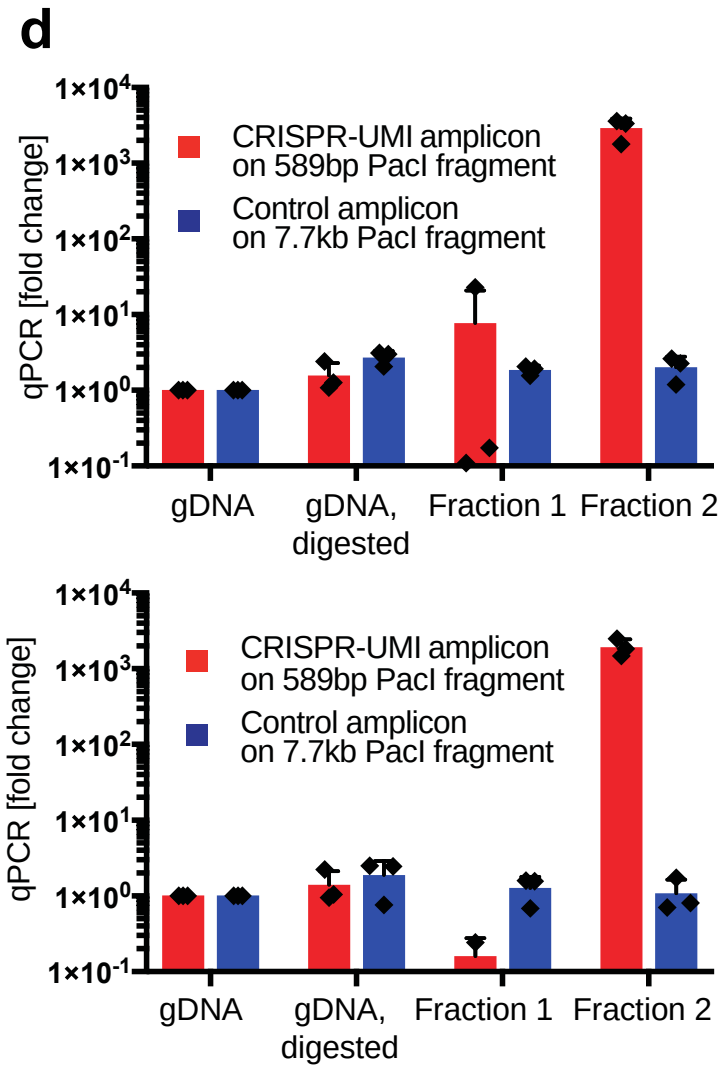
The CRISPR-UMI library is generated by 2 subsequent complex cloning steps. Initially, a random barcode consisting of 10 nucleotides is integrated into the vector backbone. Subsequently, the sgRNA pool of 26,514 sgRNAs is ligated to the barcode library with over 1000 ligation events per sgRNA. Thereby, each of the over 1000 ligation events per sgRNA combines the sgRNA with another random barcode. The combination of sgRNA and random barcodes generates a complexity of >1000 times the number of sgRNAs. We refer to this highly complex combination of sgRNA and barcode as UMI (unique molecular identifier). Our library reached a complexity of 83 million.



b

	Genes	Guides	Coverage	Complexity (million)
Transcription Factors	845	3290	1820x	6
DNA Repair	365	1437	3061x	4.4
Epigenetic regulators	548	2134	1500x	3.2
Druggable Genes	885	4453	8776x	39
Other nuclear genes 1	1994	7765	2575x	20
Other nuclear genes 2	2072	8069	954x	7.7
Handpicked	360	1417	2314x	3.2
Sum	**6560**	**26514**		**83.5**



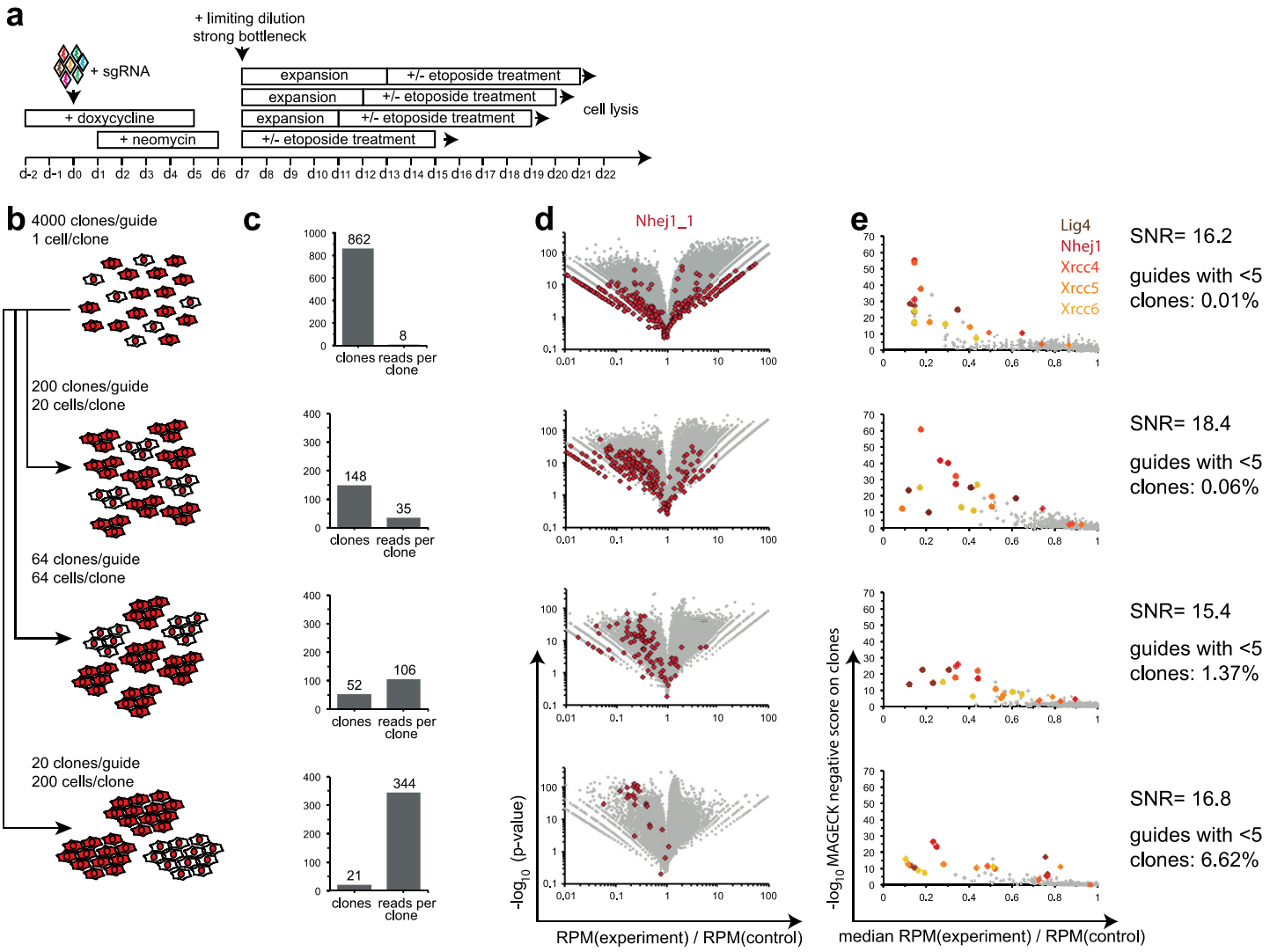


Supplementary Figure 4

CRISPR-UMI sgRNA library and cassette amplification.

(a) Vector design for library generation. Upon pooled parallel cloning of a barcode of 10 random nucleotides into retroviral backbones at complexities of 10^6, chip-synthesized sgRNA pools (at a complexity of 26514) were cloned into UMI containing backbone at a coverage >1000 clones/guide. Cassette-flanking PacI sites allowed for liberation of small sgRNA containing fragments from mammalian genomic DNA; (b) Library subpools and cloning complexity resulting in overall complexity of 83 million. (c) Ethidium bromide stained agarose gel,

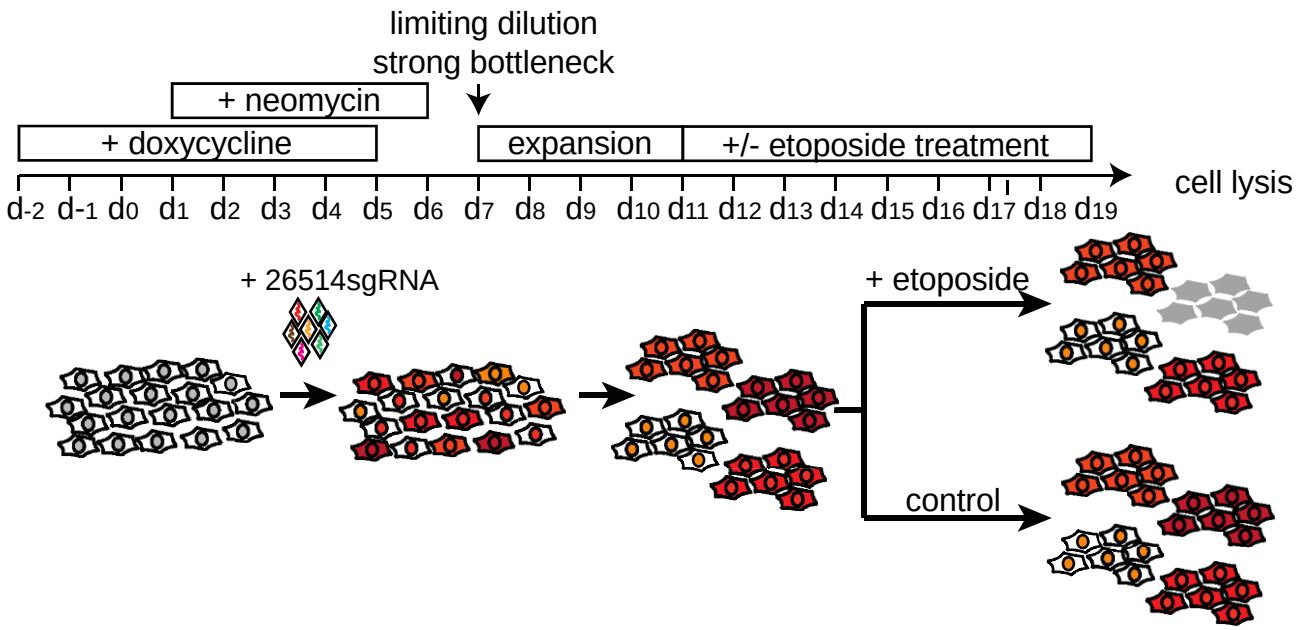
200ng DNA/lane; Digest of genomic DNA after screens and plasmid DNA as control with the octamer recognition site enzyme PacI results in mostly large genomic fragments, while sgRNA fragments are 589 bp long (arrowhead). Long and short fragments can be fractionated using magnetic beads. (d) q-PCR on genomic DNA (gDNA), PacI digest gDNA and size separated fractions of digested gDNA. Error bars are s.d., shown are two biologically independent experiments in technical triplicate, equivalent to Figure 2b, all data points are shown



Supplementary Figure 5

A pilot screen to identify optimal conditions for UMI-based CRISPR screen analysis.

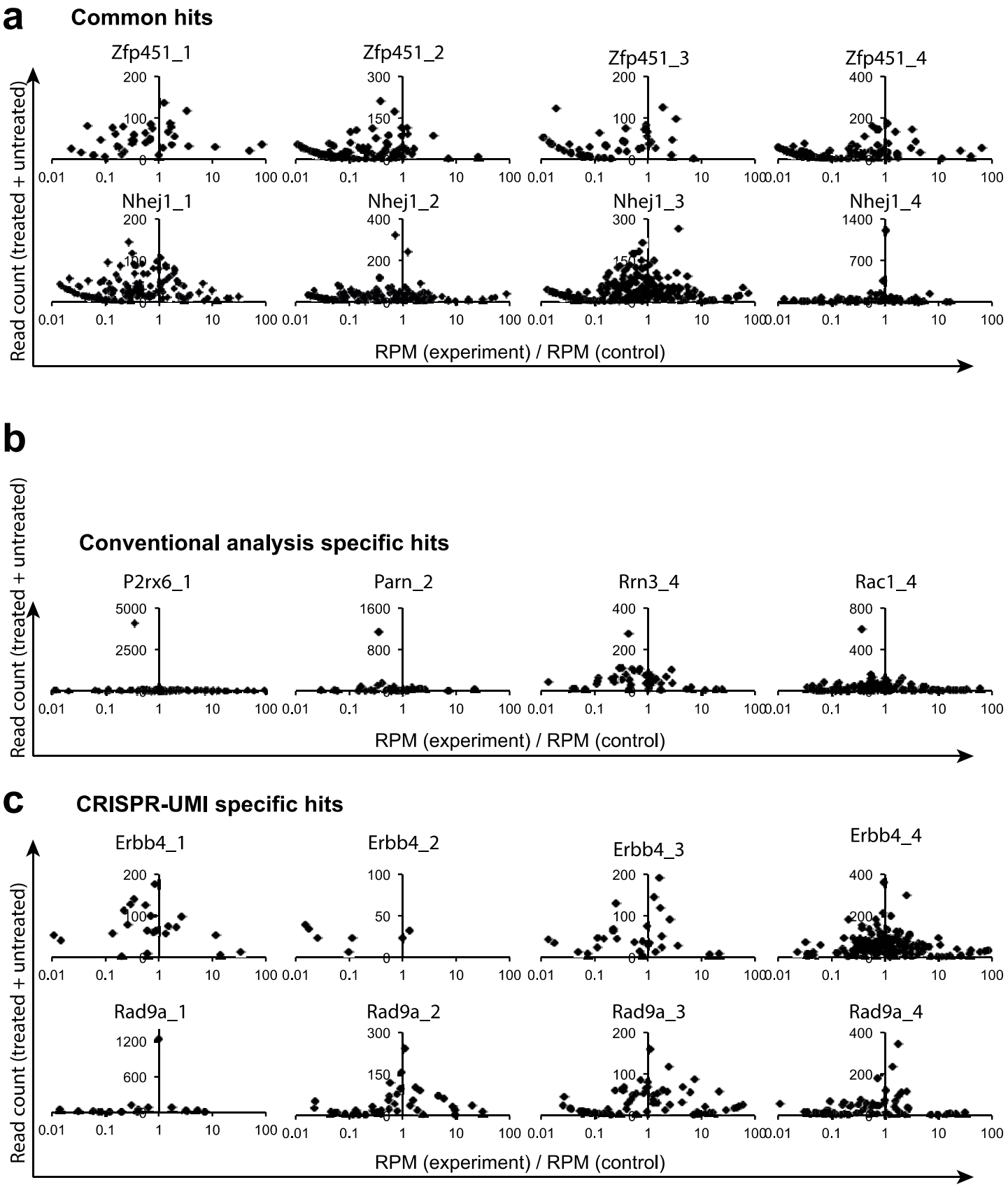
(a) Setup of screen; Upon editing, various clonal outgrowth regiments, followed by clonal expansion and dropout screening, were run in parallel. Cas9 expression was induced by Dox, selection for cells harboring guide RNAs was performed by neomycin (G418) selection. Limiting dilution and expansion is variable in the experiment. Cells are treated with or without 3.3nM etoposide a LD_{30} for 8 days. (b) Scheme illustrating variation in clone number and size (c) Average clone numbers and size determined from NGS data (d) Distribution of single cell derived clones in each regimen illustrated with guide_1 against Nhej1. P-value for each clone correlates with read depth but results in less data points. (e) Plot illustrating median dropout for each condition as well as p-value determined by combining multiple clones using MAGeCK. Signal to noise ratios (SNR) are highest in 148 clones of 35 reads, and the percentage of guides expected to have less than 5 clones due to variability in representation is with 0.06% lower than for 52 or 21 clone datasets.



Supplementary Figure 6

The screen layout for a sensitizer screen against etoposide.

Graphical illustration for large scale screen setup used to identify sensitizing mutations for etoposide. Cas9 expression was induced by doxcycline, selection for cells harboring guide RNAs was performed by neomycin (G418) selection. After washout of doxycycline and neomycin single cell derived clones are generated by a limiting dilution and clonal expansion. Cells are treated with 3.3nM Etoposide or mock treatment for 8days.

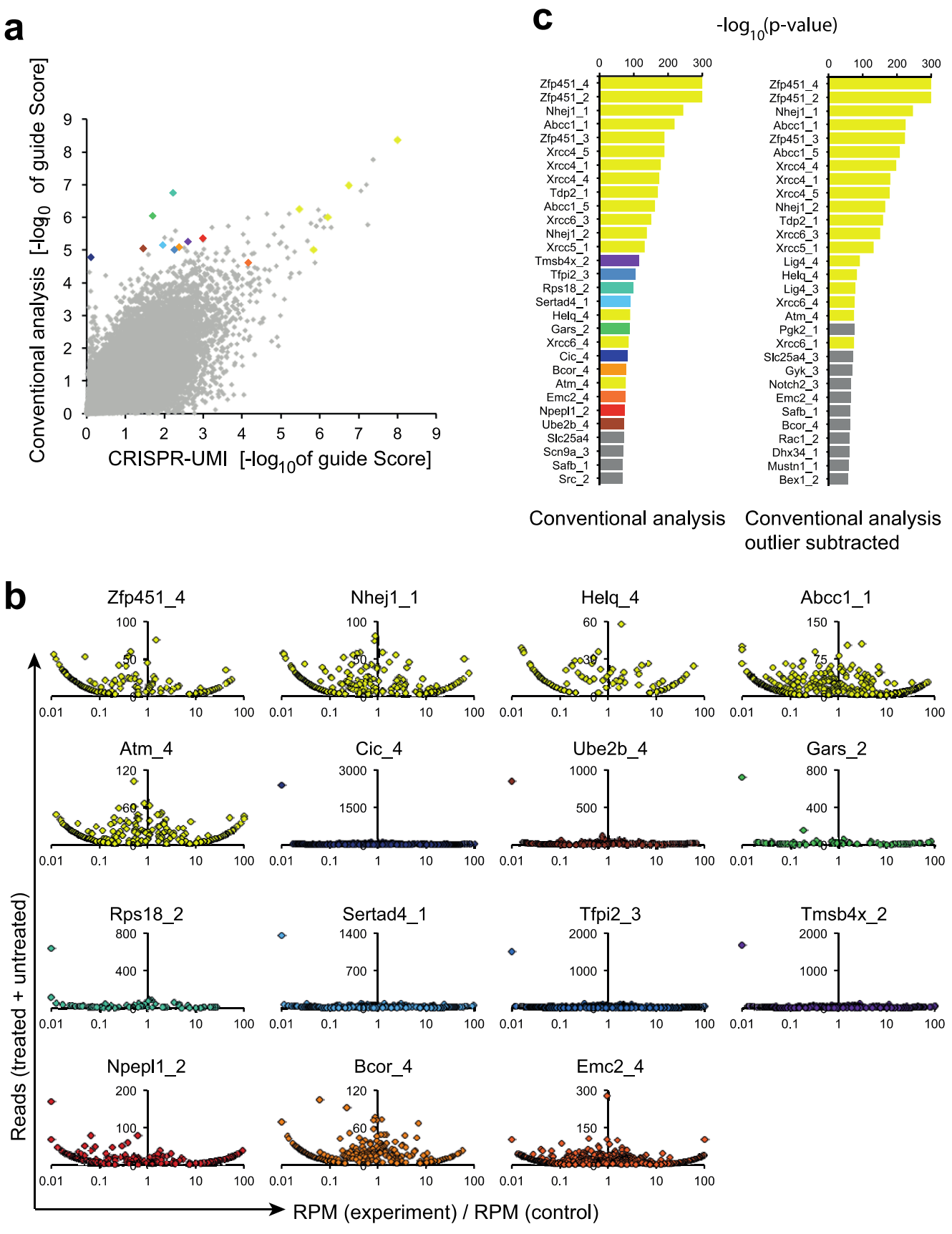


Nature Methods: doi:10.1038/nmeth.4466

Supplementary Figure 7

Read distributions of single-cell-derived clones.

(a) No strong outlier clones are detected in hits identified by conventional analysis as well as CRISPR-UMI (b) Strong outlier clones with very high read counts as well as depletion are seen in putative false positive hits called by conventional analysis. (c) Genes identified only in CRISPR-UMI show modest but reproducible depletion in multiple independent clones but are often dominated by clones with high read counts that to not deplete.

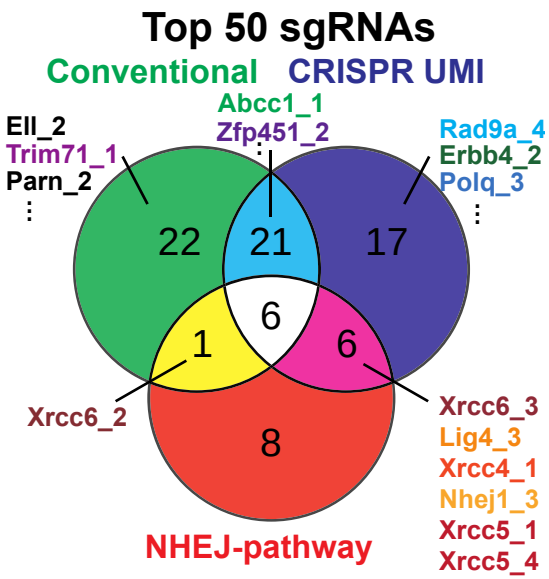


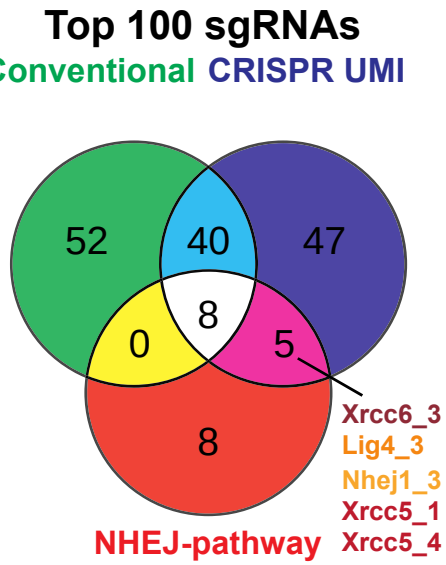
Nature Methods: doi:10.1038/nmeth.4466

Supplementary Figure 8

A pooled dropout screen without clonal outgrowth showing outliers resulting in false positive calls.

(a) Comparison of CRISPR-UMI with conventional screen analysis on guide level in absence of clonal dilution and outgrowth shows highly correlative results (yellow) as well as discrepancy between both regimen (mixed colors, Pearson correlation: 0.729) (b) While correlating sgRNAs do not contain strong outlier clones based on total read count, guides only called in conventional analysis show outlier clones responsible for overall dropout. (c) Ranking of sgRNAs improves upon removal of outlier clones (top 3 clones by read count) from the dataset illustrating their confounding effects.

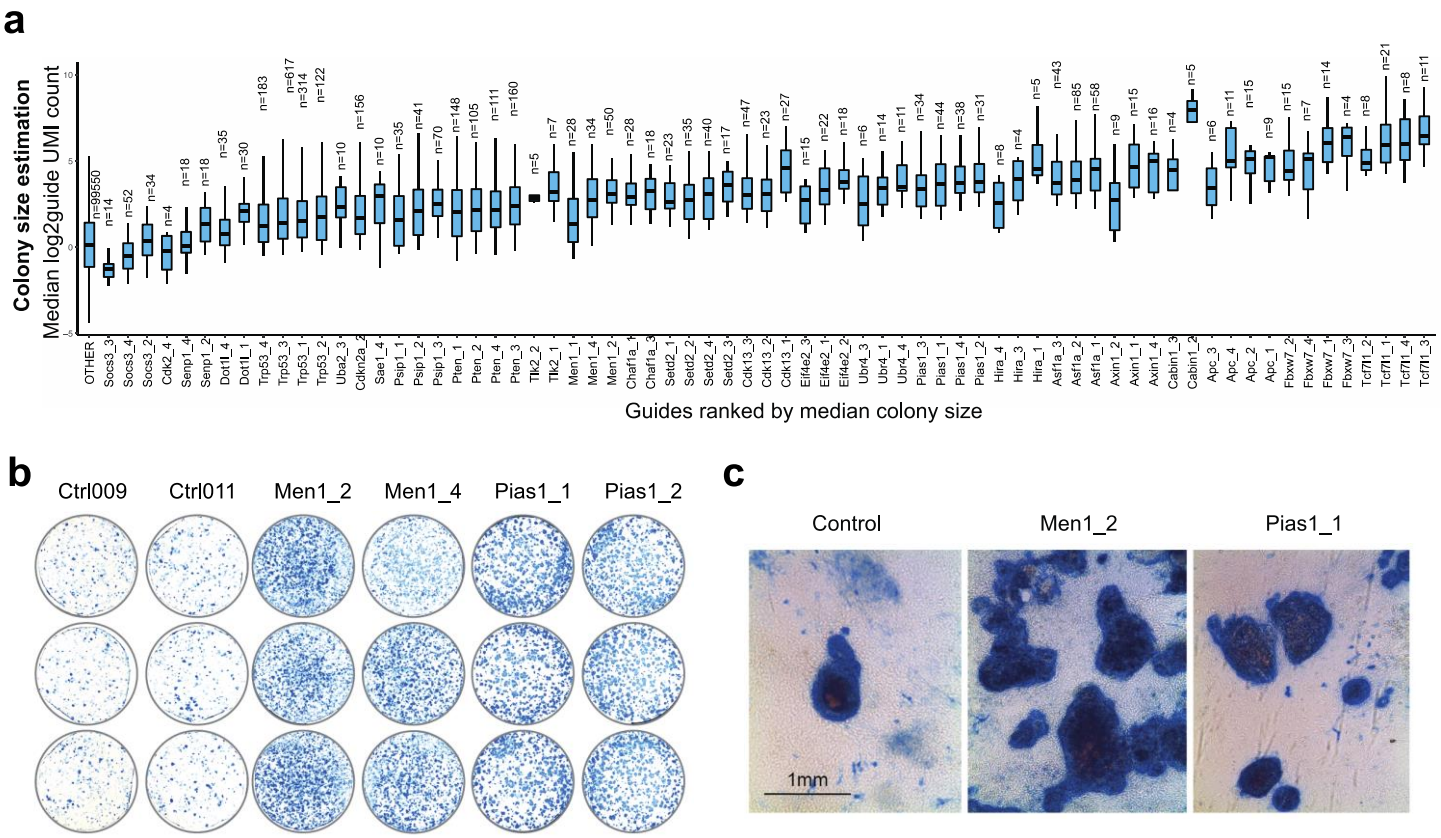




Supplementary Figure 9

A comparison of positive control guide detection in conventional analysis versus CRISPR-UMI.

Venn Diagrams illustrating the number of sgRNAs targeting the positive controls of the NHEJ complex (*Lig4, Xrcc4-6, Nhej1*) called within the top 50/100 hits.



Supplementary Figure 10

Validation of the reprogramming efficiency and predicted size distribution of colonies by UMI analysis from NGS data.

(a) Median size distribution of read counts per UMI for each sgRNA. Reads for each UMI were filtered for sequencing errors and median colony size is plotted relative to median size illustrating a marked size increase per iPS colony in many but not all identified roadblocks of reprogramming. Center line, median; hinges, 25th and 75th quartiles; whiskers,
median ± 1.58Å~ the interquartile range (IQR)/2. Individual data points represent outliers. (b) Alkaline phosphatase staining in 6 well dishes 10 days after Dox induction in the transgenic system illustrating enhanced iPS colony formation for guides targeting *Men1* or *Pias1*. (c) Representative colonies for comparison with Figure 5e stained with alkaline phosphatase in validation experiment on day 10 after Dox.

Discussion

Direct lineage conversions offer a powerful experimental framework to study cell identity, genome regulation, and TF properties. Besides their use in basic research, they offer potential in clinical applications, such as derivation of patient-specific disease models and cellular replacement therapies. However, despite significant progress, several key questions remain unresolved. Why do certain TFs succeed in reprogramming cell identity while others fail? what defines their potency and specificity? How these TFs and starting cellular context influence differentiation trajectory, overcome host cell genetic network and induces target cell type? Moreover, the paths cells take between initial and terminal states are often complex and poorly understood, with multiple distinct routes potentially existing between cell identities. Addressing these questions is essential for both refining reprogramming strategies and deepening our understanding of cell fate plasticity.

In this thesis, I compare the mechanisms underlying directed differentiation of mouse ESC to iN by ectopic expression of proneural TFs Ascl1 or Ngn2. Although the initial pluripotency and the terminal neuronal states are similar for both factors, iN conversion mechanisms differ markedly depending on whether Ascl1 or Ngn2 is expressed. First, the two factors trigger distinct transcriptional responses, activating only partially overlapping gene sets and differentially downregulating PPN genes. Second, they rely on different genetic dependencies, suggesting that each factor engages unique molecular pathways to induce neuronal identity. To dissect these dependencies, I contributed to developing CRISPR-UMI, a pooled CRISPR-Cas9 screening tool that enables assessment of clonal dynamics, reduces noise, and allows more precise gene function inference by accounting for variability in lineage transitions. Lastly, both Ascl1 and Ngn2 induces distinct alternative cell lineages, indicating divergent transcription factor mode of action and cell lineage trajectories during iN conversion.

Ascl1 promiscuity and induction of alternative lineages.

Ascl1 is a pan-neuronal TF essential for the neural stem cells differentiation into neurons during development. However, it exhibits a surprising degree of functional promiscuity when expressed outside of its native developmental context [24,117]. This promiscuity enables Ascl1 to induce alternative, non-neuronal fates, underscoring the complexity and context-dependence of TF-mediated lineage conversions.

In addition to neuronal targets, in mouse ESC Ascl1 binds and activates genes associated with the trophectoderm lineage. Indeed, Ascl1 expression leads to the formation of trophoblast-like cells (iT) marked by the expression of *Krt8* and *Cdx2*. This likely results from Ascl1 mimicking Ascl2, a bHLH TF essential for trophoblast development and the formation of placenta, as both TF share nearly identical DNA binding domains and recognize CAG(C/G)TG E-box motifs [43,118]. In turn, Ascl2 expression in ESC

also leads to the formation of both iN and iT, highlighting their overlapping transcriptional potential. Notably, ablation of de novo DNA methylation in ESC leads to the upregulation of Ascl2, directing ESC differentiation towards trophoblasts [119]. This suggests a low epigenetic barrier between pluripotent and trophoblast states, which Ascl1 manages to exploit due to promiscuous binding.

Similarly, Ascl1 overexpressed in MEFs engages muscle lineage genes, akin to muscle specific bHLH Myod1 [78]. This results in generation of myocytes, as alternative lineage to iN [78,83]. Furthermore, Ascl1 binding in MEF is also enriched for cardiac genes and coexpression with Mef2c, a common TF in induced cardiomyocyte transdifferentiation, leads to an efficient cardiomyocyte conversion [120]. Surprisingly, Ascl1 binds both myogenic and cardiac genes in ESC, as well as trophoblast specific genes in MEF, confirming Ascl1 as an "on-target" pioneering factor, that engages its targets in a wide range of cells. However, no induction of either lineages can be seen in the respective cellular context. Thus, promiscuity of Ascl1 and ability of inducing alternative lineages are at least partially constrained to the host cell lineage, likely due to expressed interaction partners or cofactors, and favorable epigenetic landscape similar to the alternative lineages.

Ascl1 promiscuity highlights the inherent plasticity and stochasticity of TF-mediated cellular reprogramming, emphasizing the need for precise control over lineage outcomes. For this conversion strategies can be refined for lineage specificity, for example by co-expression of Myt1l, a transcriptional repressor of non-neuronal programs [84] . Finally, understanding the mechanisms by which TF engage with the chromatin landscape to activate divergent programs across cellular contexts is critical for developing safer and more effective regenerative therapies.

Pluripotency shutdown dynamics of Ascl1 mediated iN conversion

A major roadblock in cellular reprogramming is overcoming the original cell identity, and its erasure is essential for establishing a new, stable fate [49,65,127] . Incomplete silencing of the starting cell program can lead to hybrid or unstable phenotypes, reduced conversion efficiency, and impaired functionality of the target cell type [79,105].

Although Ascl1 is primarily characterized as a transcriptional activator [43,78], Ascl1 rapidly downregulates key components of the PPN. Within 24 hours, Oct4 (encoded by *Pou5f1*) and *Nanog* expression is lost and *Sox2* significantly reduced by 48 hours. This is similar to the swift disassembly of the gene regulatory network observed in MEFs upon Ascl1 expression [16,42] . Interestingly, Ascl1 to iN conversion is not affected by a knockout of the transcriptional repressor Myt1l, which is induced by Ascl1 [42,84] . In contrasts, Ngn2 driven iN conversion leads to a slower PPN genes repression, suggesting distinct exit of pluripotency strategies and mode of action between Ascl1 and Ngn2. However, the precise molecular mechanisms underlying Ascl1-induced repression of pluripotency remain to be elucidated.

Development of CRIPSR-UMI and mechanistic insights into iN conversion.

Cellular reprogramming is inherently noisy due to variability in initiation times and conversion rates of individual cells, as well as differences in cell cycle between cell states – all of which can bias experimental outcomes. To counteract variability of lineage transitions, we developed CRISPR-UMI, a barcoded lineage-tracing CRISPR screening platform. By tagging each sgRNA with a unique molecular identifier (UMI), CRISPR-UMI allows tracking of individual clones, thereby capturing heterogeneity in reprogramming dynamics as well as noise reduction due to outlier discrimination. Most importantly, it offers an additional layer of insight into the screen data without altering with conventional CRISPR screening workflows.

We first applied CRISPR-UMI in a pooled forward genetic screen to identify roadblocks in MEF to iPSC reprogramming. The screen successfully identified known reprogramming roadblocks such as *Trp53*, *Pten*, *Chaf1a*, *Dot1l*, *Socs3*, and *Uba2* [65,121–124] , validating the results. Beyond confirming the expected hits, CRISPR-UMI provided insight into the dynamics of colony formation, revealing functional differences between phenotypes based on colony size and number. For example, sgRNAs targeting *Axin1*, *Apc*, and *Tcf7l1*, key components of Wnt signaling as well as negative regulators of PPN, produced few but large colonies, suggesting efficient reprogramming with strong proliferative capacity. In contrast, targeting *Dot1l*, *Socs3*, and *Senp1* resulted in many small colonies, indicating enhanced reprogramming initiation but limited expansion. CRISPR-UMI thus enables a refined understanding of reprogramming heterogeneity by linking clonal behavior to genetic perturbations, which would be missed in conventional analyses.

Additionally, screen identified *Pias1*, an E3 SUMO-protein ligase, and Menin, encoded by the gene *Men1*, neither of which was previously implicated in iPSC reprogramming. Pias1 promotes SUMOylation of p53 and c-Jun, leading to activation of pro-apoptosis related genes like Bax[125] . As reprogramming induces cellular stress and DNA damage response, knockout of Pias1 could act as reduce cell cycle arrest or apoptosis, increasing reprogramming efficiency. Menin acts as a nuclear scaffold protein, interacting with chromatin modifiers and transcription factors to regulate gene expression, and thus involved in both lineage identity maintenance and differentiation[126] . Interestingly, Menin has dual roles in cancer biology showing both tumor suppressor and oncogenic functions in different tumors[126] . For example, in neuroendocrine tumors, Menin suppresses tumorigenesis by regulating of cell cycle inhibitors such as p18 and p27[126] . This suggests that Menin could pose a dual barrier for iPSC reprogramming for stabilizing MEF state and limiting activation of cell cycle by OKSM factors.

In summary, CRISPR-UMI positive-selection screens can distinguish between genes affecting the number of reprogramming events and those influencing colony growth rate, revealing both known and novel barriers to iPSC generation. CRISPR-UMI also detects clonal effects, reducing false positives arising from double infections and allows higher multiplicity of infection. This improves reproducibility and allows for smaller or more complex experiments, making it a powerful tool for studying stochastic events like differentiation and metastasis.

I then conducted comparative forward genetic screens using CRISPR-UMI to systematically elucidate the different genetic dependencies of Ascl1 and Ngn2-induced transitions. The results revealed that Ascl1-driven iN conversion is particularly sensitive to regulators of the PPN and the cell cycle, including *Tcf7l1*, *Apc*, *Ptpn2*, *Ccnk*, and *Cdk13*. Notably, inhibition of these factors abolished iN formation but not the generation of iTs, suggesting that repression of PPN and cell cycle arrest are specifically critical for Ascl1 mediated neuronal conversion, controlling balance between iN and iT.

Tcf7l1 is a key transcriptional repressor in ESC, where it plays a crucial role in maintaining the balance between pluripotency and differentiation[128–130] . Its activity is tightly regulated, and its overexpression promotes exit from the pluripotent state, making it a critical factor in early developmental decisions [128–130] . Interestingly, although Ascl1 depends on Tcf7l1 at the onset of iN conversion, pluripotency genes were still repressed upon knockout of Tcf7l1. However, in the absence of Tcf7l1, Ascl1 fails to induce the expression of *Cdkn1c*, a cyclin-dependent kinase inhibitor that controls cell proliferation, promotes cell cycle exit, and facilitates differentiation [131] . This suggests that Tcf7l1 may be necessary for establishing a permissive chromatin environment for Cdkn1c activation. Importantly, overexpression of Cdkn1c with Ascl1 restores cell cycle exit and rescues the neuronal conversion in the absence of Tcf7l1. Furthermore, coexpression of Cdkn1c with Ascl1 improved the maturity of both iN and iT phenotypes, suggesting that enforced cell cycle exit can be an improvement in designing new iN strategies.

Altogether, Ascl1's ability to efficiently downregulate the initial cell GRN, together with efficient cell cycle arrest, likely contributes to its ability to drive iN conversion across large developmental distances. By simultaneously silencing the original cell identity and halting proliferation, fewer cues would be required to establish a new cellular state. However, this also creates permissive conditions in which promiscuous binding of Ascl1 may steer cells towards alternative fates.

Robust transcriptional network activation and induction of intermediate states by Ngn2

Ngn2 is a well-established proneural TF that has been widely used in iN conversion protocols [17,19,44,45,116] . However, functional differences and redundancies are less understood compared to Ascl1, limiting the rational selection of TF for reprogramming strategies.

Ngn2 shows a preference for binding to a distinct subset of E-box motifs from Ascl1 with consensus sequence CAGATG. This results in only partially overlapping transcriptional output between Ascl1 and Ngn2. These differences can be attributed to the DNA-binding domain (DBD) rather than the activation domain or cofactor interactions [43,78] . Notably, replacing Ascl1 DBD with that of Ngn2 was shown to be sufficient to redirect Ascl1 to Ngn2 binding sites in mouse ESC [43,78] .

Ngn2 induces a more robust and more interconnected transcriptional network, including activation of other potent proneural TF such as *Neurod1* and *Neurog1*. Additionally, Ngn2 binds and induces a set of genes associated with NSC identity, such as *Pax3* and *Notch1*, leading to the formation of an NSC-like proliferative intermediate. Despite this, Ngn2 shows slower and weaker repression of the PPN, compared to Ascl1. This is evidenced by the formation of trophoblast- or primitive endoderm-like cells following acute knockout of *Nanog* or *Pou5f1*, respectively. These findings suggest that residual PPN activity persists during Ngn2-driven iN conversion, promoting the formation of the side lineages [132–134] . Curiously, although Ngn2 engages neuronal genes in MEF, it is insufficient on its own to transdifferentiate MEF to iN [45,46] . However, coexpression of Ngn2 together with Myt1l enables iN conversion, suggesting that weaker repression of the starting cell identity may restrict Ngn2 transdifferentiating potential [13] .

Similar to other PPN factors, Sox2 is not downregulated and remains expressed in the NSC-like cells. As Sox2 is a hallmark and a driver of NSC identity, it appears to be repurposed from the PPN into NSC gene regulatory network. However, the establishment of a stable NSC state is not essential for iN generation, as *Sox2* knockout does not impair neuronal conversion but does eliminate the NSC-like cells.

In summary, Ngn2 establishes a more stable and self-sustaining transcriptional network than Ascl1. Its ability to repurpose components of the PPN and generate intermediate NSC-like state providing a more efficient trajectory explaining faster and more efficient iN conversion.

Conclusions and outlook

This study highlights the complexity and context-dependence of TF-mediated lineage conversions, using Ascl1 and Ngn2 iN directed differentiation from ESC as a model system. Despite their shared ability to generate iN, Ascl1 and Ngn2 do so through distinct molecular mechanisms. Ascl1 rapidly dismantles ESC network, followed by establishment of neuronal identity. However, Ascl1 exhibits high functional promiscuity by directly binding and inducing genes of non-neuronal lineages in a context dependent manner. In contrast, Ngn2 induces a more robust and self-sustaining transcriptional network which is overlayed over the ESC network in a developmental-like progression.

Nonetheless, several questions remain and should be a focus in future research. Promiscuity of TFs like Ascl1 and how to effectively negate will be an important challenge for therapeutic applications. The use of mutant TFs for cellular reprogramming is an avenue remains to be explored. For example, TF properties such as phosphorylation can significantly influence conversion efficiency and specificity [77,135,136] . Phosphorylation of Ascl1 and Ngn2 by cell cycle kinases are less neurogenic and use of phosphor-mutants can increase activation of neuronal maturation genes and more efficient repression of host cell identity [77,135,136] . An intriguing direction for future work is development of chimeric or TFs with mutagenized DBD, which could be optimized for high neuronal conversion efficiency with reduced binding promiscuity[43,118].

Addition of the transcriptional repressors e.g. Myt1l, as well as inhibition of key alternative lineage genes could lead to improvements in reprogramming protocols [53,55,84,117]. Alternatively, non-coding RNAs, such as miR-9/9* and miR-124, have been shown to promote neuronal identity by repressing non-neuronal gene networks and enhancing chromatin accessibility at neuronal loci [137,138] . Similarly, shRNA mediated knockdown of lineage-specific repressors, for example neuronal lineage repressor REST, can facilitate neuronal conversions [137,138] .

Activity of TFs is also influenced by interaction partners, and their addition to the reprogramming TF cocktails could tailor the cellular outcome. For example, E47 enhances Ascl1-mediated neurogenesis, but reduces Ngn2 induced neurogenesis in chick embryos [139] . Furthermore, Neurog2 homodimers are more neurogenic than Ascl1 homodimers or Ngn2-E47 heterodimers [139,140] . Lastly, interaction partners could also divert TFs towards different or partial binding motifs changing transcriptional output toward target lineages [141,142] .

Beyond individual TF behavior, a more detailed mechanistic understanding is lacking. Lineage conversions are a multistep process with downstream targets further facilitating the conversion [16,49,50,70,75] . Understanding how TFs engage downstream targets to steer lineage trajectories and outcompete the host cell program will be essential for refining conversion strategies while suppressing side lineages, which could be illuminated by single-cell multi-omics [19,83,89] . Furthermore, comparing reprogramming regimes that use similar TFs across different cell types can reveal general principles as well as context-specific constraints. For example, contrasting ESC to iN and MEF to iN conversions with the same TF would clarify how TFs engage and activate lineage-specific programs among different epigenetic landscapes and could help to define the minimal and essential requirements for successful iN induction. Identifying essential genes and pathways required for successful conversion can further help refine reprogramming cocktails and reduce variability in outcomes.

Lastly, as Ascl1 and, to lesser extent, Ngn2 are implicated in oncogenesis such as glioblastoma and non-small lung cancers [143–145] . Thus, understanding their mode of action in disease versus reprogramming is essential for developing safer regenerative therapies.

Ultimately, these insights can be leveraged to design safer and more efficient reprogramming protocols for regenerative medicine, enabling the generation of patient-specific cell types for disease modeling, drug screening, and cellular therapies.

References

1. Holmberg, J., and Perlmann, T. (2012). Maintaining differentiated cellular identity. Nat Rev Genet *13*, 429–439. https://doi.org/10.1038/nrg3209.

2. Gurdon, J.B. (1962). The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles. Development *10*, 622–640. https://doi.org/10.1242/dev.10.4.622.

3. Principles of Cloning (2014). (Elsevier).

4. Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H.S. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810–813. https://doi.org/10.1038/385810a0.

5. Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.-S., Sritanaudomchai, H., et al. (2013). Human Embryonic Stem Cells Derived by Somatic Cell Nuclear Transfer. Cell *153*, 1228–1238. https://doi.org/10.1016/j.cell.2013.05.006.

6. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

7. Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., et al. (2023). A transcription factor atlas of directed differentiation. Cell *186*, 209-229.e26. https://doi.org/10.1016/j.cell.2022.11.026.

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

9. 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. https://doi.org/10.1016/j.cell.2006.07.024.

10. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872. https://doi.org/10.1016/j.cell.2007.11.019.

11. Nefzger, C.M., Rossello, F.J., Chen, J., Liu, X., Knaupp, A.S., Firas, J., Paynter, J.M., Pflueger, J., Buckberry, S., Lim, S.M., et al. (2017). Cell Type of Origin Dictates the Route to Pluripotency. Cell Reports *21*, 2649–2660. https://doi.org/10.1016/j.celrep.2017.11.029.

12. Zhao, X., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., Guo, C., Ma, Q., Wang, L., et al. (2009). iPS cells produce viable mice through tetraploid complementation. Nature *461*, 86–90. https://doi.org/10.1038/nature08267.

13. Chanda, S., Ang, C.E., Davila, J., Pak, C., Mall, M., Lee, Q.Y., Ahlenius, H., Jung, S.W., Südhof, T.C., and Wernig, M. (2014). Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. Stem Cell Reports *3*, 282–296. https://doi.org/10.1016/j.stemcr.2014.05.020.

14. 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. https://doi.org/10.1038/nature08797.

15. Marro, S., Pang, Z.P., Yang, N., Tsai, M.-C., Qu, K., Chang, H.Y., Südhof, T.C., and Wernig, M. (2011). Direct Lineage Conversion of Terminally Differentiated Hepatocytes to Functional Neurons. Cell Stem Cell *9*, 374–382. https://doi.org/10.1016/j.stem.2011.09.002.

16. 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. https://doi.org/10.1016/j.cell.2013.09.028.

17. Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., et al. (2013). Rapid Single-Step Induction of Functional Neurons from Human Pluripotent Stem Cells. Neuron *78*, 785–798. https://doi.org/10.1016/j.neuron.2013.05.029.

18. Briggs, J.A., Li, V.C., Lee, S., Woolf, C.J., Klein, A., and Kirschner, M.W. (2017). Mouse embryonic stem cells can differentiate via multiple paths to the same state. eLife *6*, e26945. https://doi.org/10.7554/elife.26945.

19. Lin, H.-C., He, Z., Ebert, S., Schörnig, M., Santel, M., Nikolova, M.T., Weigert, A., Hevers, W., Kasri, N.N., Taverna, E., et al. (2021). NGN2 induces diverse neuron types from human pluripotency. Stem Cell Rep. *16*, 2118–2127. https://doi.org/10.1016/j.stemcr.2021.07.006.

20. Liu, G., David, B.T., Trawczynski, M., and Fessler, R.G. (2020). Advances in Pluripotent Stem Cells: History, Mechanisms, Technologies, and Applications. Stem Cell Rev and Rep *16*, 3–32. https://doi.org/10.1007/s12015-019-09935-x.

21. Fan, D., Wu, H., Pan, K., Peng, H., and Wu, R. (2021). Regenerating Damaged Myocardium: A Review of Stem-Cell Therapies for Heart Failure. Cells *10*, 3125. https://doi.org/10.3390/cells10113125.

22. Vasan, L., Park, E., David, L.A., Fleming, T., and Schuurmans, C. (2021). Direct Neuronal Reprogramming: Bridging the Gap Between Basic Science and Clinical Application. Front. Cell Dev. Biol. *9*, 681087. https://doi.org/10.3389/fcell.2021.681087.

23. Huyghe, A., Furlan, G., Schroeder, J., Cascales, E., Trajkova, A., Ruel, M., Stüder, F., Larcombe, M., Sun, Y.B.Y., Mugnier, F., et al. (2022). Comparative roadmaps of reprogramming and oncogenic transformation identify Bcl11b and Atoh8 as broad regulators of cellular plasticity. Nat Cell Biol *24*, 1350–1363. https://doi.org/10.1038/s41556-022-00986-w.

24. Wilkinson, G., Dennis, D., and Schuurmans, C. (2013). Proneural genes in neocortical development. Neuroscience *253*, 256–273. https://doi.org/10.1016/j.neuroscience.2013.08.029.

25. Shimojo, H., Ohtsuka, T., and Kageyama, R. (2008). Oscillations in Notch Signaling Regulate Maintenance of Neural Progenitors. Neuron *58*, 52–64. https://doi.org/10.1016/j.neuron.2008.02.014.

26. Imayoshi, I., Isomura, A., Harima, Y., Kawaguchi, K., Kori, H., Miyachi, H., Fujiwara, T., Ishidate, F., and Kageyama, R. (2013). Oscillatory Control of Factors Determining Multipotency and Fate in Mouse Neural Progenitors. Science *342*, 1203–1208. https://doi.org/10.1126/science.1242366.

27. Imayoshi, I., and Kageyama, R. (2014). bHLH Factors in Self-Renewal, Multipotency, and Fate Choice of Neural Progenitor Cells. Neuron *82*, 9–23. https://doi.org/10.1016/j.neuron.2014.03.018.

28. Kageyama, R., Ohtsuka, T., Shimojo, H., and Imayoshi, I. (2008). Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. Nat. Neurosci. *11*, 1247–1251. https://doi.org/10.1038/nn.2208.

29. Castro, D.S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L.G., Hunt, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. Genes Dev. *25*, 930–945. https://doi.org/10.1101/gad.627811.

30. 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. https://doi.org/10.1101/gad.940902.

31. 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. https://doi.org/10.1038/sj.emboj.7600278.

32. 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. https://doi.org/10.1101/gad.14.1.67.

33. Andersen, J., Urbán, N., Achimastou, A., Ito, A., Simic, M., Ullom, K., Martynoga, B., Lebel, M., Göritz, C., Frisén, J., et al. (2014). A Transcriptional Mechanism Integrating Inputs from Extracellular Signals to Activate Hippocampal Stem Cells. Neuron *83*, 1085–1097. https://doi.org/10.1016/j.neuron.2014.08.004.

34. Urbán, N., Berg, D.L.C. van den, Forget, A., Andersen, J., Demmers, J.A.A., Hunt, C., Ayrault, O., and Guillemot, F. (2016). Return to quiescence of mouse neural stem cells by degradation of a proactivation protein. Science *353*, 292–295. https://doi.org/10.1126/science.aaf4802.

35. Perino, M., and Veenstra, G.J.C. (2016). Chromatin Control of Developmental Dynamics and Plasticity. Dev. Cell *38*, 610–620. https://doi.org/10.1016/j.devcel.2016.08.004.

36. Henikoff, S., and Greally, J.M. (2016). Epigenetics, cellular memory and gene regulation. Curr. Biol. *26*, R644–R648. https://doi.org/10.1016/j.cub.2016.06.011.

37. Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678. https://doi.org/10.1038/nrg.2016.112.

38. Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. Cancer Discovery *12*, 31–46. https://doi.org/10.1158/2159-8290.cd-21-1059.

39. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.

40. Farsetti, A., Illi, B., and Gaetano, C. (2023). How epigenetics impacts on human diseases. Eur. J. Intern. Med. *114*, 15–22. https://doi.org/10.1016/j.ejim.2023.05.036.

41. Hepburn, A.C., Steele, R.E., Veeratterapillay, R., Wilson, L., Kounatidou, E.E., Barnard, A., Berry, P., Cassidy, J.R., Moad, M., El-Sherif, A., et al. (2019). The induction of core pluripotency master regulators in cancers defines poor clinical outcomes and treatment resistance. Oncogene *38*, 4412–4424. https://doi.org/10.1038/s41388-019-0712-y.

42. Wapinski, O.L., Lee, Q.Y., Chen, A.C., Li, R., Corces, M.R., Ang, C.E., Treutlein, B., Xiang, C., Baubet, V., Suchy, F.P., et al. (2017). Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. Cell Reports *20*, 3236–3247. https://doi.org/10.1016/j.celrep.2017.09.011.

43. 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. https://doi.org/10.1038/s41593-019-0399-y.

44. 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. https://doi.org/10.1016/j.stem.2016.11.006.

45. 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. https://doi.org/10.1038/ncomms3183.

46. 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. https://doi.org/10.1016/j.stemcr.2016.09.013.

47. Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. Trends in Genetics *32*, 29–41. https://doi.org/10.1016/j.tig.2015.11.001.

48. Nicetto, D., and Zaret, K.S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Current Opinion in Genetics & Development *55*, 1–10. https://doi.org/10.1016/j.gde.2019.04.013.

49. Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. Cell *151*, 994–1004. https://doi.org/10.1016/j.cell.2012.09.045.

50. Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. Cell *150*, 1209–1222. https://doi.org/10.1016/j.cell.2012.08.023.

51. Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Mancarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602. https://doi.org/10.1038/nature10953.

52. Becker, J.S., McCarthy, R.L., Sidoli, S., Donahue, G., Kaeding, K.E., He, Z., Lin, S., Garcia, B.A., and Zaret, K.S. (2017). Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. Molecular Cell *68*, 1023-1037.e15. https://doi.org/10.1016/j.molcel.2017.11.030.

53. Lim, B., Domsch, K., Mall, M., and Lohmann, I. (2024). Canalizing cell fate by transcriptional repression. Mol. Syst. Biol. *20*, 144–161. https://doi.org/10.1038/s44320-024-00014-z.

54. Schoenherr, C.J., and Anderson, D.J. (1995). The Neuron-Restrictive Silencer Factor (NRSF): A Coordinate Repressor of Multiple Neuron-Specific Genes. Science *267*, 1360–1363. https://doi.org/10.1126/science.7871435.

55. 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. https://doi.org/10.1016/j.stem.2015.05.014.

56. Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49–55. https://doi.org/10.1038/nature07056.

57. 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. https://doi.org/10.1038/nbt1374.

58. Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2008). c-Myc Is Dispensable for Direct Reprogramming of Mouse Fibroblasts. Cell Stem Cell *2*, 10–12. https://doi.org/10.1016/j.stem.2007.12.001.

59. Ernst, C. (2016). Proliferation and Differentiation Deficits are a Major Convergence Point for Neurodevelopmental Disorders. Trends Neurosci. *39*, 290–299. https://doi.org/10.1016/j.tins.2016.03.001.

60. Homem, C.C.F., Repic, M., and Knoblich, J.A. (2015). Proliferation control in neural stem and progenitor cells. Nat. Rev. Neurosci. *16*, 647–659. https://doi.org/10.1038/nrn4021.

61. Fishman, V., Shnayder, T., Orishchenko, K., Bader, M., Alenina, N., and Serov, O. (2015). Cell divisions are not essential for the direct conversion of fibroblasts into neuronal cells. Cell Cycle *14*, 1188–1196. https://doi.org/10.1080/15384101.2015.1012875.

62. Jiang, H., Xu, Z., Zhong, P., Ren, Y., Liang, G., Schilling, H.A., Hu, Z., Zhang, Y., Wang, X., Chen, S., et al. (2015). Cell cycle and p53 gate the direct conversion of human fibroblasts to dopaminergic neurons. Nat Commun *6*, 10100. https://doi.org/10.1038/ncomms10100.

63. 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. https://doi.org/10.7554/elife.41356.

64. Budhavarapu, V.N., Chavez, M., and Tyler, J.K. (2013). How is epigenetic information maintained through DNA replication? Epigenetics & Chromatin *6*, 32. https://doi.org/10.1186/1756-8935-6-32.

65. Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Ang, C.E., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. Nature *528*, 218–224. https://doi.org/10.1038/nature15749.

66. Liu, Y., Pelham-Webb, B., Giammartino, D.C.D., Li, J., Kim, D., Kita, K., Saiz, N., Garg, V., Doane, A., Giannakakou, P., et al. (2017). Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. Cell Reports *19*, 1283–1293. https://doi.org/10.1016/j.celrep.2017.04.067.

67. Consortium, Roadmap Epigenomics, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.

68. Zaret, K.S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. Annu. Rev. Genet. *54*, 367–385. https://doi.org/10.1146/annurev-genet-030220-015007.

69. Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. Cell *161*, 555–568. https://doi.org/10.1016/j.cell.2015.03.017.

70. 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. https://doi.org/10.1016/j.stem.2016.11.006.

71. Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., and Zaret, K.S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev. *10*, 1670–1682. https://doi.org/10.1101/gad.10.13.1670.

72. Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature *475*, 386–389. https://doi.org/10.1038/nature10116.

73. Kogiso, T., Nagahara, H., Otsuka, M., Shiratori, K., and Dowdy, S.F. (2013). Transdifferentiation of human fibroblasts into hepatocyte-like cells by defined transcriptional factors. Hepatol Int *7*, 937–944. https://doi.org/10.1007/s12072-013-9432-5.

74. Rombaut, M., Boeckmans, J., Rodrigues, R.M., Grunsven, L.A. van, Vanhaecke, T., and Kock, J.D. (2021). Direct reprogramming of somatic cells into induced hepatocytes: Cracking the Enigma code. Journal of Hepatology *75*, 690–705. https://doi.org/10.1016/j.jhep.2021.04.048.

75. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459.e20. https://doi.org/10.1016/j.cell.2016.12.016.

76. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. Cell Stem Cell *14*, 838–853. https://doi.org/10.1016/j.stem.2014.04.003.

77. Woods, L.M., Ali, F.R., Gomez, R., Chernukhin, I., Marcos, D., Parkinson, L.M., Tayoun, A.N.A., Carroll, J.S., and Philpott, A. (2022). Elevated ASCL1 activity creates de novo regulatory elements associated with neuronal differentiation. BMC Genomics *23*, 255. https://doi.org/10.1186/s12864-022-08495-8.

78. 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. https://doi.org/10.1038/s41556-020-0490-3.

79. Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. Cell *176*, 928-943.e22. https://doi.org/10.1016/j.cell.2019.01.006.

80. Camacho-Aguilar, E., Warmflash, A., and Rand, D.A. (2021). Quantifying cell transitions in C. elegans with data-fitted landscape models. PLoS Comput. Biol. *17*, e1009034. https://doi.org/10.1371/journal.pcbi.1009034.

81. Sáez, M., Blassberg, R., Camacho-Aguilar, E., Siggia, E.D., Rand, D.A., and Briscoe, J. (2022). Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. Cell Syst. *13*, 12-28.e3. https://doi.org/10.1016/j.cels.2021.08.013.

82. Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. Phys. Rev. Lett. *94*, 128701. https://doi.org/10.1103/physrevlett.94.128701.

83. Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395. https://doi.org/10.1038/nature18323.

84. Mall, M., Kareta, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake, S., Ang, C.E., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. Nature *544*, 245–249. https://doi.org/10.1038/nature21722.

85. Ruiz, S., Panopoulos, A.D., Herrerías, A., Bissig, K.-D., Lutz, M., Berggren, W.T., Verma, I.M., and Belmonte, J.C.I. (2011). A High Proliferation Rate Is Required for Cell Reprogramming and Maintenance of Human Embryonic Stem Cell Identity. Curr. Biol. *21*, 45–52. https://doi.org/10.1016/j.cub.2010.11.049.

86. Hanna, J., Saha, K., Pando, B., Zon, J. van, Lengner, C.J., Creyghton, M.P., Oudenaarden, A. van, and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601. https://doi.org/10.1038/nature08592.

87. Francesconi, M., Stefano, B.D., Berenguer, C., Andrés-Aguayo, L. de, Plana-Carmona, M., Mendez-Lago, M., Guillaumet-Adkins, A., Rodriguez-Esteban, G., Gut, M., Gut, I.G., et al. (2019). Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming. eLife *8*, e41627. https://doi.org/10.7554/elife.41627.

88. Balsalobre, A., and Drouin, J. (2022). Pioneer factors as master regulators of the epigenome and cell fate. Nat Rev Mol Cell Biol. https://doi.org/10.1038/s41580-022-00464-z.

89. Stuart, H.T., Stirparo, G.G., Lohoff, T., Bates, L.E., Kinoshita, M., Lim, C.Y., Sousa, E.J., Maskalenka, K., Radzisheuskaya, A., Malcolm, A.A., et al. (2019). Distinct Molecular Trajectories Converge to Induce Naive Pluripotency. Cell Stem Cell *25*, 388-406.e8. https://doi.org/10.1016/j.stem.2019.07.009.

90. Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324. https://doi.org/10.1038/nature05944.

91. 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. https://doi.org/10.1126/science.1151526.

92. Barrachina, L., Arshaghi, T.E., O'Brien, A., Ivanovska, A., and Barry, F. (2023). Induced pluripotent stem cells in companion animals: how can we move the field forward? Front. Vet. Sci. *10*, 1176772. https://doi.org/10.3389/fvets.2023.1176772.

93. Lee, S.-G., Mikhalchenko, A.E., Yim, S.H., Lobanov, A.V., Park, J.-K., Choi, K.-H., Bronson, R.T., Lee, C.-K., Park, T.J., and Gladyshev, V.N. (2017). Naked Mole Rat Induced Pluripotent Stem Cells and Their Contribution to Interspecific Chimera. Stem Cell Reports *9*, 1706–1720. https://doi.org/10.1016/j.stemcr.2017.09.013.

94. Tan, L., Ke, Z., Tombline, G., Macoretta, N., Hayes, K., Tian, X., Lv, R., Ablaeva, J., Gilbert, M., Bhanu, N.V., et al. (2017). Naked Mole Rat Cells Have a Stable Epigenome that Resists iPSC Reprogramming. Stem Cell Reports *9*, 1721–1734. https://doi.org/10.1016/j.stemcr.2017.10.001.

95. Miyawaki, S., Kawamura, Y., Oiwa, Y., Shimizu, A., Hachiya, T., Bono, H., Koya, I., Okada, Y., Kimura, T., Tsuchiya, Y., et al. (2016). Tumour resistance in induced pluripotent stem cells derived from naked mole-rats. Nat Commun *7*, 11471. https://doi.org/10.1038/ncomms11471.

96. Simpson, D.J., Olova, N.N., and Chandra, T. (2021). Cellular reprogramming and epigenetic rejuvenation. Clin. Epigenetics *13*, 170. https://doi.org/10.1186/s13148-021-01158-7.

97. Lapasset, L., Milhavet, O., Prieur, A., Besnard, E., Babled, A., Aït-Hamou, N., Leschik, J., Pellestor, F., Ramirez, J.-M., Vos, J.D., et al. (2011). Rejuvenating senescent and centenarian human cells by reprogramming through the pluripotent state. Genes Dev. *25*, 2248–2253. https://doi.org/10.1101/gad.173922.111.

98. Huh, C.J., Zhang, B., Victor, M.B., Dahiya, S., Batista, L.F., Horvath, S., and Yoo, A.S. (2016). Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts. eLife *5*, e18648. https://doi.org/10.7554/elife.18648.

99. Mertens, J., Paquola, A.C.M., Ku, M., Hatch, E., Böhnke, L., Ladjevardi, S., McGrath, S., Campbell, B., Lee, H., Herdy, J.R., et al. (2015). Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. Cell Stem Cell *17*, 705–718. https://doi.org/10.1016/j.stem.2015.09.001.

100. Yang, Y., Jiao, J., Gao, R., Le, R., Kou, X., Zhao, Y., Wang, H., Gao, S., and Wang, Y. (2015). Enhanced Rejuvenation in Induced Pluripotent Stem Cell-Derived Neurons Compared with Directly Converted Neurons from an Aged Mouse. Stem Cells and Development *24*, 2767–2777. https://doi.org/10.1089/scd.2015.0137.

101. Parmar, M., Torper, O., and Drouin-Ouellet, J. (2019). Cell-based therapy for Parkinson's disease: A journey through decades toward the light side of the Force. Eur J Neurosci *49*, 463–471. https://doi.org/10.1111/ejn.14109.

102. 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. https://doi.org/10.1016/j.stem.2013.12.001.

103. Colarusso, J.L., and Zhou, Q. (2022). Direct Reprogramming of Different Cell Lineages into Pancreatic β-Like Cells. Cell. Reprogram. *24*, 252–258. https://doi.org/10.1089/cell.2022.0048.

104. Xiao, X., Guo, P., Shiota, C., Zhang, T., Coudriet, G.M., Fischbach, S., Prasadan, K., Fusco, J., Ramachandran, S., Witkowski, P., et al. (2018). Endogenous Reprogramming of Alpha Cells into Beta Cells, Induced by Viral Gene Therapy, Reverses Autoimmune Diabetes. Cell Stem Cell *22*, 78-90.e4. https://doi.org/10.1016/j.stem.2017.11.020.

105. Furuyama, K., Chera, S., Gurp, L. van, Oropeza, D., Ghila, L., Damond, N., Vethe, H., Paulo, J.A., Joosten, A.M., Berney, T., et al. (2019). Diabetes relief in mice by glucose-sensing insulin-secreting human α-cells. Nature *567*, 43–48. https://doi.org/10.1038/s41586-019-0942-8.

106. Lee, H., Lee, H.Y., Lee, B.E., Gerovska, D., Park, S.Y., Zaehres, H., Araúzo-Bravo, M.J., Kim, J.-I., Ha, Y., Schöler, H.R., et al. (2020). Sequentially induced motor neurons from human fibroblasts facilitate locomotor recovery in a rodent spinal cord injury model. eLife *9*, e52069. https://doi.org/10.7554/elife.52069.

107. Su, Z., Niu, W., Liu, M.-L., Zou, Y., and Zhang, C.-L. (2014). In vivo conversion of astrocytes to neurons in the injured adult spinal cord. Nat. Commun. *5*, 3338. https://doi.org/10.1038/ncomms4338.

108. Gourdie, R.G., Dimmeler, S., and Kohl, P. (2016). Novel therapeutic strategies targeting fibroblasts and fibrosis in heart disease. Nat Rev Drug Discov *15*, 620–638. https://doi.org/10.1038/nrd.2016.89.

109. Qian, L., Huang, Y., Spencer, C.I., Foley, A., Vedantham, V., Liu, L., Conway, S.J., Fu, J., and Srivastava, D. (2012). In vivo reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature *485*, 593–598. https://doi.org/10.1038/nature11044.

110. Song, K., Nam, Y.-J., Luo, X., Qi, X., Tan, W., Huang, G.N., Acharya, A., Smith, C.L., Tallquist, M.D., Neilson, E.G., et al. (2012). Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature *485*, 599–604. https://doi.org/10.1038/nature11139.

111. Gutierrez-Aranda, I., Ramos-Mejia, V., Bueno, C., Munoz-Lopez, M., Real, Pedro.J., Mácia, A., Sanchez, L., Ligero, G., Garcia-Parez, J.L., and Menendez, P. (2010). Human Induced Pluripotent Stem Cells Develop Teratoma More Efficiently and Faster Than Human Embryonic Stem Cells Regardless the Site of Injection. Stem Cells *28*, 1568–1570. https://doi.org/10.1002/stem.471.

112. 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. https://doi.org/10.1016/j.stem.2015.12.003.

113. Yang, N., Chanda, S., Marro, S., Ng, Y.-H., Janas, J.A., Haag, D., Ang, C.E., Tang, Y., Flores, Q., Mall, M., et al. (2017). Generation of pure GABAergic neurons by transcription factor programming. Nat Methods *14*, 621–628. https://doi.org/10.1038/nmeth.4291.

114. Black, J.B., McCutcheon, S.R., Dube, S., Barrera, A., Klann, T.S., Rice, G.A., Adkar, S.S., Soderling, S.H., Reddy, T.E., and Gersbach, C.A. (2020). Master Regulators and Cofactors of Human Neuronal Cell Fate Specification Identified by CRISPR Gene Activation Screens. Cell Reports *33*, 108460. https://doi.org/10.1016/j.celrep.2020.108460.

115. 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. https://doi.org/10.1126/sciadv.abe8978.

116. Hulme, A.J., Maksour, S., Glover, M.S.-C., Miellet, S., and Dottori, M. (2022). Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. Stem Cell Reports *17*, 14–34. https://doi.org/10.1016/j.stemcr.2021.11.015.

117. 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. https://doi.org/10.1038/s41556-020-0490-3.

118. Casey, B.H., Kollipara, R.K., Pozo, K., and Johnson, J.E. (2018). Intrinsic DNA binding properties demonstrated for lineage-specifying basic helix-loop-helix transcription factors. Genome Res. *28*, 484–496. https://doi.org/10.1101/gr.224360.117.

119. Kinoshita, M., Li, M.A., Barber, M., Mansfield, W., Dietmann, S., and Smith, A. (2021). Disabling de novo DNA methylation in embryonic stem cells allows an illegitimate fate trajectory. Proc. Natl. Acad. Sci. U.S.A. *118*, e2109475118. https://doi.org/10.1073/pnas.2109475118.

120. Wang, H., Keepers, B., Qian, Y., Xie, Y., Colon, M., Liu, J., and Qian, L. (2022). Cross-lineage potential of Ascl1 uncovered by comparing diverse reprogramming regulatomes. Cell Stem Cell *29*, 1491-1504.e9. https://doi.org/10.1016/j.stem.2022.09.006.

121. Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Mancarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602. https://doi.org/10.1038/nature10953.

122. Liao, J., Marumoto, T., Yamaguchi, S., Okano, S., Takeda, N., Sakamoto, C., Kawano, H., Nii, T., Miyamoto, S., Nagai, Y., et al. (2013). Inhibition of PTEN Tumor Suppressor Promotes the Generation of Induced Pluripotent Stem Cells. Mol. Ther. *21*, 1242–1250. https://doi.org/10.1038/mt.2013.60.

123. Marión, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature *460*, 1149–1153. https://doi.org/10.1038/nature08287.

124. Buckley, S.M., Aranda-Orgilles, B., Strikoudis, A., Apostolou, E., Loizou, E., Moran-Crusio, K., Farnsworth, C.L., Koller, A.A., Dasgupta, R., Silva, J.C., et al. (2012). Regulation of Pluripotency and Cellular Reprogramming by the Ubiquitin-Proteasome System. Cell Stem Cell *11*, 783–798. https://doi.org/10.1016/j.stem.2012.09.011.

125. Schmidt, D., and Müller, S. (2002). Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity. Proc. Natl. Acad. Sci. *99*, 2872–2877. https://doi.org/10.1073/pnas.052559499.

126. Majer, A.D., Hua, X., and Katona, B.W. (2024). Menin in Cancer. Genes *15*, 1231. https://doi.org/10.3390/genes15091231.

127. Wille, C.K., Zhang, X., Haws, S.A., Denu, J.M., and Sridharan, R. (2023). DOT1L is a barrier to histone acetylation during reprogramming to pluripotency. Sci. Adv. *9*, eadf3980. https://doi.org/10.1126/sciadv.adf3980.

128. Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev. *22*, 746–755. https://doi.org/10.1101/gad.1642408.

129. Pereira, L., Yi, F., and Merrill, B.J. (2006). Repression of Nanog Gene Transcription by Tcf3 Limits Embryonic Stem Cell Self-Renewal. Mol Cell Biol *26*, 7479–7491. https://doi.org/10.1128/mcb.00368-06.

130. Yi, F., Pereira, L., Hoffman, J.A., Shy, B.R., Yuen, C.M., Liu, D.R., and Merrill, B.J. (2011). Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nat Cell Biol *13*, 762–770. https://doi.org/10.1038/ncb2283.

131. Creff, J., and Besson, A. (2020). Functional Versatility of the CDK Inhibitor p57Kip2. Front. Cell Dev. Biol. *8*, 584590. https://doi.org/10.3389/fcell.2020.584590.

132. Bates, L.E., Alves, M.R.P., and Silva, J.C.R. (2021). Auxin-degron system identifies immediate mechanisms of OCT4. Stem Cell Reports *16*, 1818–1831. https://doi.org/10.1016/j.stemcr.2021.05.016.

133. Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372–376. https://doi.org/10.1038/74199.

134. Hough, S.R., Clements, I., Welch, P.J., and Wiederholt, K.A. (2006). Differentiation of Mouse Embryonic Stem Cells after RNA Interference-Mediated Silencing of OCT4 and Nanog. Stem Cells *24*, 1467–1475. https://doi.org/10.1634/stemcells.2005-0475.

135. Azzarelli, R., McNally, A., Dell'Amico, C., Onorati, M., Simons, B., and Philpott, A. (2022). ASCL1 phosphorylation and ID2 upregulation are roadblocks to glioblastoma stem cell differentiation. Sci Rep *12*, 2341. https://doi.org/10.1038/s41598-022-06248-x.

136. 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. https://doi.org/10.1242/dev.106377.

137. Drouin-Ouellet, J., Lau, S., Brattås, P.L., Ottosson, D.R., Pircs, K., Grassi, D.A., Collins, L.M., Vuono, R., Sjöland, A.A., Westergren-Thorsson, G., et al. (2017). REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways. EMBO Mol. Med. *9*, 1117–1131. https://doi.org/10.15252/emmm.201607471.

138. Lu, Y.-L., and Yoo, A.S. (2018). Mechanistic Insights Into MicroRNA-Induced Neuronal Reprogramming of Human Adult Fibroblasts. Front. Neurosci. *12*, 522. https://doi.org/10.3389/fnins.2018.00522.

139. Dréau, G.L., Escalona, R., Fueyo, R., Herrera, A., Martínez, J.D., Usieto, S., Menendez, A., Pons, S., Martinez-Balbas, M.A., and Marti, E. (2018). E proteins sharpen neurogenesis by modulating proneural bHLH transcription factors' activity in an E-box-dependent manner. eLife *7*, e37267. https://doi.org/10.7554/elife.37267.

140. Li, S., Mattar, P., Zinyk, D., Singh, K., Chaturvedi, C.-P., Kovach, C., Dixit, R., Kurrasch, D.M., Ma, Y.-C., Chan, J.A., et al. (2012). GSK3 Temporally Regulates Neurogenin 2 Proneural Activity in the Neocortex. J. Neurosci. *32*, 7791–7805. https://doi.org/10.1523/jneurosci.1309-12.2012.

141. Heidt, A.B., Rojas, A., Harris, I.S., and Black, B.L. (2007). Determinants of Myogenic Specificity within MyoD Are Required for Noncanonical E Box Binding. Mol. Cell. Biol. *27*, 5910–5920. https://doi.org/10.1128/mcb.01700-06.

142. Molkentin, J.D., Black, B.L., Martin, J.F., and Olson, E.N. (1995). Cooperative activation of muscle gene expression by MEF2 and myogenic bHLH proteins. Cell *83*, 1125–1136. https://doi.org/10.1016/0092-8674(95)90139-6.

143. Su, Z., Zang, T., Liu, M.-L., Wang, L.-L., Niu, W., and Zhang, C.-L. (2014). Reprogramming the fate of human glioma cells to impede brain tumor development. Cell Death Dis. *5*, e1463–e1463. https://doi.org/10.1038/cddis.2014.425.

144. Augustyn, A., Borromeo, M., Wang, T., Fujimoto, J., Shao, C., Dospoy, P.D., Lee, V., Tan, C., Sullivan, J.P., Larsen, J.E., et al. (2014). ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. Proc. Natl. Acad. Sci. *111*, 14788–14793. https://doi.org/10.1073/pnas.1410419111.

145. Borromeo, M.D., Savage, T.K., Kollipara, R.K., He, M., Augustyn, A., Osborne, J.K., Girard, L., Minna, J.D., Gazdar, A.F., Cobb, M.H., et al. (2016). ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep. *16*, 1259–1272. https://doi.org/10.1016/j.celrep.2016.06.081.