RESEARCH ARTICLE SUMMARY

CELL REPROGRAMMING

Cell competition during reprogramming gives rise to dominant clones

Nika Shakiba*, Ahmed Fahmy*, Gowtham Jayakumaran, Sophie McGibbon, Laurent David, Daniel Trcka, Judith Elbaz, Mira C. Puri, Andras Nagy, Derek van der Kooy, Sidhartha Goyal, Jeffrey L. Wrana, Peter W. Zandstra†

INTRODUCTION: Recent advances in lineage tracking using DNA-based barcoding technologies have provided new insights into the developmental dynamics of cell populations. Nevertheless, the mechanisms that govern these dynamics and their relation to clonal dominance remain unclear. Understanding the factors that determine the appearance and disappearance of individual clones and their derivatives in dynamic, heterogeneous systems has broad implications for normal and aberrant tissue development, providing a new lens with which to understand emergent behaviors in multicellular systems.

The process of reprogramming, whereby somatic cells are converted to induced pluripotent stem cells (iPSCs) through the overexpression of key transcription factors, serves as a tractable model to probe the connection between intrinsic and extrinsic factors influencing cell population dynamics. The field of reprogramming has largely accepted the "clonal equipotency" paradigm, in which somatic cells have an equal potential to attain an iPSC state; however, it remains unclear whether inequalities between the fitness of reprogramming cells drive non-neutral clonal drift.

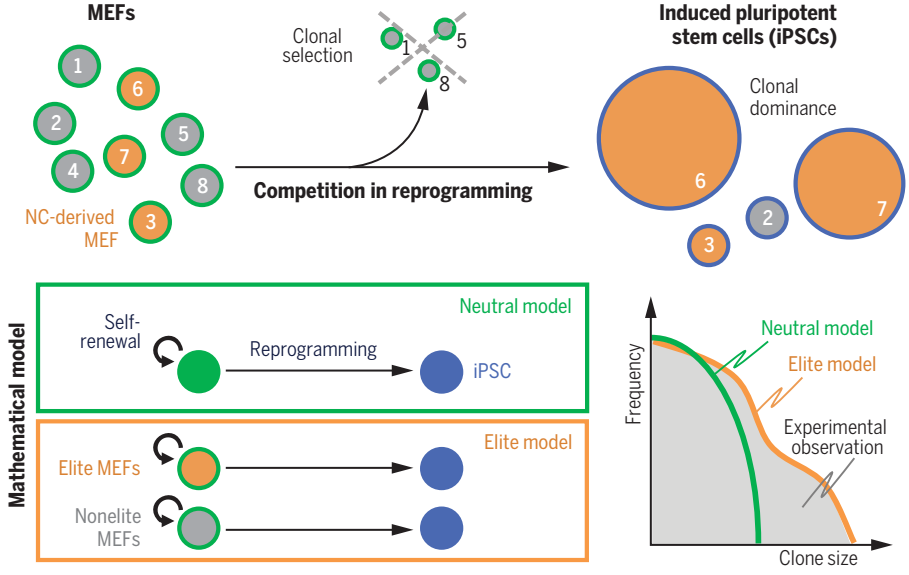
RATIONALE: This study aimed to quantitatively explore the impact of competitive interactions in driving clonal dynamics within a reprogramming population. A cellular barcoding strategy, whereby heritable DNA tags are inserted into the genome of mouse embryonic fibroblasts (MEFs), was used to track the relative contribution of clonal derivatives to the reprogramming population. Mathematical modeling was used to decouple the effect of stochastic reprogramming latencies on clonal dynamics, enabling the identification of non-neutral clonal dominance. Finally, a mouse model was developed to trace the developmental origins of the heterogeneous MEF compartment and investigate the mechanism of cell fitness inequalities in reprogramming.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan0925

RESULTS: Population reprogramming promoted highly selective clonal dynamics, leading to the elimination of up to 80% of clones after a week of reprogramming. Single-cell reprogramming studies, on the other hand, suggested that all MEF-derived clones had a propensity to reprogram when cultured in isolation. Furthermore, an elite subset of dominating clones was found to drive the dynamics and composition of the bulk reprogramming pool. Early dominating clones that emerged after a week of reprogramming exhibited a robust and predictable propensity to overtake the culture and contribute to the final successfully reprogrammed iPSC pool. Our modeling results reveal that neutral effects did not capture the evolution of clone size distributions. Indeed, parallel reprogramming cultures grown from a common pool of barcoded MEFs demonstrated a correlation in clonal outcomes, driven by a subset of reprogramming MEFs. This subset of MEFs exhibited an a priori propensity for reprogramming and dominance. Lineage tracing revealed that these MEFs arise during embryonic development from a Wnt1-expressing population associable with the neural crest compartment.

CONCLUSION: This study reveals cell competition as an important parameter that arises in the population context as a result of genetically encoded inequalities in cell fitness. Competition serves as a mechanism by which otherwise hidden cells with context-specific eliteness emerge to occupy and dominate the reprogramming niche. Given the propensity for a small subset of clones to overtake the population, it is likely that bulk measurements of the reprogramming process represent a biased view of the reprogramming trajectory. The findings of this study reveal that cell competition may be a generalizable and controllable parameter for understanding and controlling the dynamics of multicellular developmental and disease systems in a quantified manner. ■



Competition during reprogramming results in dominant clones derived from elite cells. (Top) During reprogramming, when MEFs transition to pluripotency, competition between cells leads to clonal selection and dominance. Numbered circles represent individual reprogramming clones; circle size indicates the relative abundance of cells derived from a clone in the overall population. (Bottom) Mathematical modeling reveals that clonal dynamics are not driven by stochastic reprogramming latencies, but rather by a subset of elite MEFs derived from the neural crest (NC) that have an enhanced a priori reprogramming propensity.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: peter.zandstra@ubc.ca
Cite this article as N. Shakiba *et al.*, *Science* **364**, eaan0925 (2019). DOI: 10.1126/science.aan0925

RESEARCH ARTICLE

CELL REPROGRAMMING

Cell competition during reprogramming gives rise to dominant clones

Nika Shakiba[1]*, Ahmed Fahmy[2]*, Gowtham Jayakumaran[2,3]†, Sophie McGibbon[4], Laurent David[5,6,7], Daniel Trcka[3], Judith Elbaz[3], Mira C. Puri[3,8], Andras Nagy[3,9,10,11], Derek van der Kooy[2,12], Sidhartha Goyal[1,4], Jeffrey L. Wrana[2,3], Peter W. Zandstra[1,12,13,14,15]‡

The ability to generate induced pluripotent stem cells from differentiated cell types has enabled researchers to engineer cell states. Although studies have identified molecular networks that reprogram cells to pluripotency, the cellular dynamics of these processes remain poorly understood. Here, by combining cellular barcoding, mathematical modeling, and lineage tracing approaches, we demonstrate that reprogramming dynamics in heterogeneous populations are driven by dominant "elite" clones. Clones arise a priori from a population of poised mouse embryonic fibroblasts derived from Wnt1-expressing cells that may represent a neural crest–derived population. This work highlights the importance of cellular dynamics in fate programming outcomes and uncovers cell competition as a mechanism by which cells with eliteness emerge to occupy and dominate the reprogramming niche.

The ability to engineer gene regulatory networks to control cell fate holds great potential for enabling regenerative medicine technologies. Perhaps the most prominent example to date is the discovery that somatic cells can be reprogrammed to generate induced pluripotent stem cells (iPSCs) by the overexpression of four key transcription factors—Oct4, Klf4, c-Myc, and Sox2 (OKMS) (*1*). Patient-specific iPSCs serve as an unlimited source of cells that give rise to all cell types in the body, with applications in tissue engineering and drug screening. Additionally, reprogramming has enhanced our insight into the nature of cellular plasticity.

Our understanding of reprogramming comes largely from population-level analyses of the cell transcriptome, proteome, and epigenome (*2*–*5*). These analyses have revealed three distinct and well-characterized phases of bulk reprogramming: Cells exit a somatic state, transition through a transgene-dependent state (pre-iPSCs), and finally acquire a stable transgene-independent state (iPSCs) (*2*–*4*). These studies do not, however, provide insight into individual cell trajectories during reprogramming (*5*–*7*). Indeed, the relationship between cell population outcomes and single-cell reprogramming events is poorly understood. Analyses of single isolated cells undergoing reprogramming (also known as clones) (*8*, *9*) support the concept of "clonal equipotency," wherein all cells are able to reprogram, albeit with stochastic reprogramming latencies (*8*) (fig. S1A). As a result, individual clones exhibiting asynchronous state changes are expected to give rise to heterogeneity in population reprogramming. The apparent contradiction between the progression of reprogramming through deterministic, stepwise phases at the population level and the stochastic nature of clonal reprogramming has yet to be resolved.

Clonal interactions give rise to both active competition, during which cells may directly induce apoptosis in their neighbors (*10*, *11*), as well as passive competition for limited resources and space (*12*). Understanding the role that cell competition plays during reprogramming is critical to interpreting the abundance of population omics data being generated, yet to date, it has not been investigated.

Here we present data on clonal competition in a population context, revealing how reprogramming can occur in a stepwise manner in spite of the stochastic nature of clonal reprogramming. We quantified heterogeneity in clonal fitness—the relative survival potential of a reprogramming clone—using barcode-based cell tracking in population reprogramming. In the context of reprogramming, fitness is a function of a cell's reprogramming potential as well as its ability to outcompete other clones. By using cell competition as a new lens, we dissect the link between clonal "fitness" and reprogramming "eliteness." Although eliteness has been used to describe clones with enhanced reprogramming potential (*8*), we expand this definition to describe clones that are dominant (where dominance is quantified by the number of cells in that clone) in the context of large heterogeneous populations. In particular, our data allowed us to examine whether clonal dynamics during reprogramming are governed by stochastic reprogramming latencies of equipotent clones or by selective clonal loss or dominance of heterogeneous clones (Fig. 1A). Given the role of cell competition in PSC culture (*13*, *14*), we hypothesized that clonal contributions to the reprogramming pool would not be a result of neutral dynamics. Additionally, we explored whether cellular heterogeneity is an existing property of somatic cells or emerges during reprogramming.

Population homogeneity emerges in spite of clonal heterogeneity

In a preliminary study, we assessed the degree to which single-cell and population reprogramming dynamics may differ. We used a secondary reprogramming system with somatic cells that contain a doxycycline (DOX)–inducible cassette to drive exogenous OKMS factor expression. Briefly, primary iPSCs (harboring DOX-inducible OKMS transgenes) were used to generate secondary mice via tetraploid complementation, herein referred to as the 1B tetr system (*15*). The resultant mice contain secondary mouse embryonic fibroblasts (MEFs) that are readily induced to reprogram via DOX induction. We used our previously reported marker strategy (*6*) to track reprogramming in these MEFs, in terms of the contribution of isolated clones to pre-iPSC (transgene-dependent, $CD24^{high}$/SSEA1+) and iPSC (transgene-independent, $CD24^{low}$/SSEA1+) states on day 18 (fig. S1B). Although clones in bulk cultures gave rise to a homogeneous population of pre-iPSCs, as previously reported (*5*, *6*), we observed a high degree of clone-to-clone heterogeneity in the pre-iPSC fraction produced by individual isolated clones (Fig. 1B).

We next determined whether the high fraction of pre-iPSCs previously seen in population reprogramming (*6*) could be recapitulated in clonal mixing studies, combining distinct clones with

[1]Institute of Biomaterials and Biomedical Engineering (IBBME), University of Toronto, Toronto, Ontario M5S 3E1, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [4]Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada. [5]SFR-SANTE, iPSC Core Facility, INSERM, CNRS, UNIV Nantes, CHU Nantes, Nantes, France. [6]CRTI, INSERM, Université de Nantes, Nantes, France. [7]ITUN, CHU Nantes, Nantes, France. [8]Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5T 3H7, Canada. [9]Department of Obstetrics and Gynecology, University of Toronto, Toronto, Ontario M5G 1E2, Canada. [10]Institute of Medical Science, University of Toronto, Toronto, Ontario M5T 3H7, Canada. [11]Australian Regenerative Medicine Institute, Monash University, Melbourne, Victoria, Australia. [12]The Donnelly Centre for Cellular and Biomolecular Research (CCBR), University of Toronto, Toronto, Ontario M5S 3E1, Canada. [13]School of Biomedical Engineering, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. [14]The Biomedical Research Centre, The University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada. [15]Michael Smith Laboratories, The University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.
*These authors contributed equally to this work.
†Present address: Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
‡Corresponding author. Email: peter.zandstra@ubc.ca

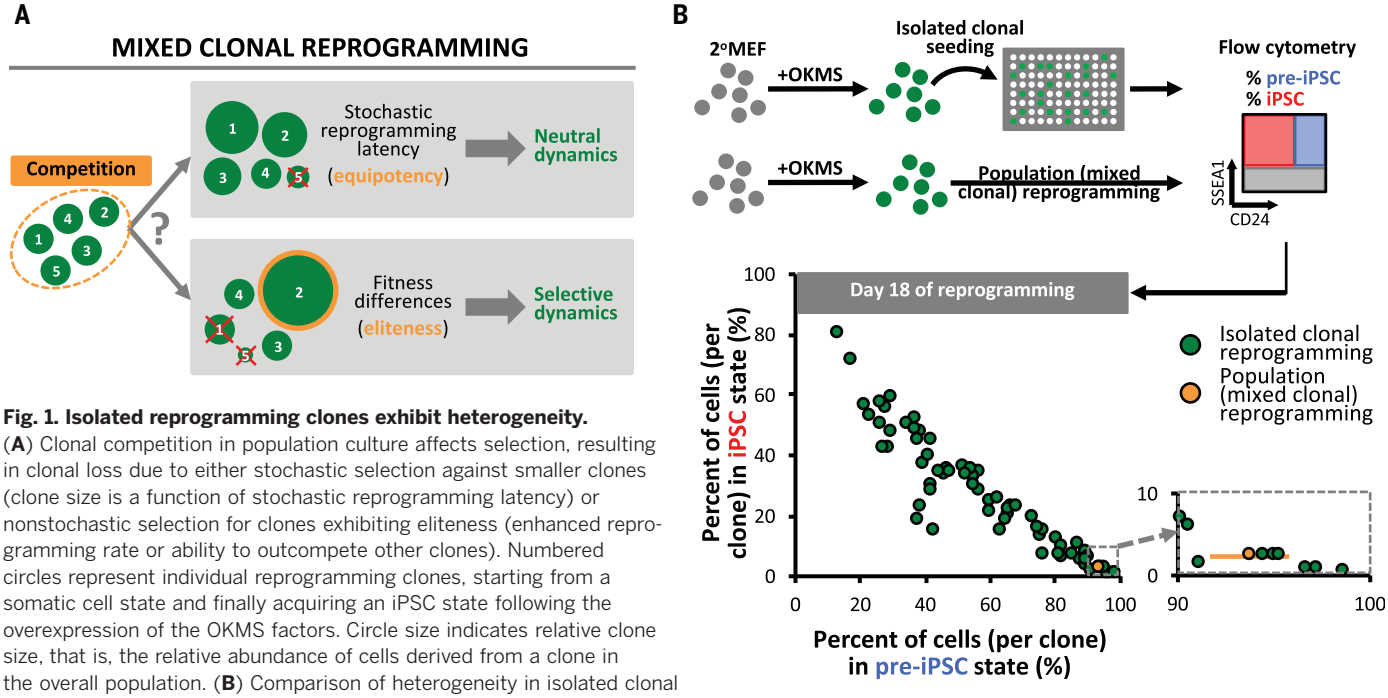


Fig. 1. Isolated reprogramming clones exhibit heterogeneity. (**A**) Clonal competition in population culture affects selection, resulting in clonal loss due to either stochastic selection against smaller clones (clone size is a function of stochastic reprogramming latency) or nonstochastic selection for clones exhibiting eliteness (enhanced reprogramming rate or ability to outcompete other clones). Numbered circles represent individual reprogramming clones, starting from a somatic cell state and finally acquiring an iPSC state following the overexpression of the OKMS factors. Circle size indicates relative clone size, that is, the relative abundance of cells derived from a clone in the overall population. (**B**) Comparison of heterogeneity in isolated clonal versus population reprogramming. Ability of individual secondary reprogramming MEFs (2°MEFs) versus population MEF reprogramming (1B tetr secondary reprogramming system) to give rise to pre-iPSCs (CD24^{high}/SSEA1+) and iPSCs (CD24^{low}/SSEA1+) is shown, assayed by flow cytometry at day 18 (D18) of reprogramming. Reprogramming clones were derived from DOX-induced secondary MEFs that were single-cell sorted on D8. Population reprogramming data represent means ± standard deviation of n = 3 biological replicates.

varying fractions of pre-iPSCs (fig. S1C). Homogeneous pre-iPSC populations emerge from cocultures of two or more heterogeneous clones (fig. S1, D and E). This finding revealed the possibility that a subset of clones could drive the population's dynamics and motivated further studies to understand the underlying clonal competition during reprogramming.

Cellular barcoding strategy reveals clonal selection during reprogramming

To probe long-term clonal dynamics over the 30-day reprogramming process, we adapted a previously reported cellular barcoding strategy (*16*). In brief, before inducing reprogramming, a variable DNA sequence—or "barcode"—was integrated into the MEFs' genome by lentiviral transduction (Fig. 2A). To ensure that our observations were not an artifact of the system and represented the reprogramming process, we conducted experiments with four reprogramming systems containing DOX-inducible OKMS factor gene cassettes: the secondary 1B tetr system (*15*), the secondary iRep1 system [derived by previous methods (*15*)], a primary reprogramming system (figs. S2 and S3), and the secondary 9DT system [similar to that previously reported (*17*), fig. S4].

Barcoded MEFs from the four systems were sampled for initial barcode profiling before inducing reprogramming. Throughout the reprogramming process, cells were sampled regularly to assess barcode abundance at time points based on pluripotency characterization studies either published in the 1B tetr system (*5*) or conducted here (figs. S2 to S4). At day 14, half of the cells in the 1B tetr and iRep1 systems were transitioned to DOX-free (DOX−) culture to assess clonal contributions to transgene-independent iPSCs (*5*) (Fig. 2A). A detailed description of the experimental and bioinformatics pipeline can be found in the methods section and figs. S5 and S6.

To assess clonal selection, the number of clones present in the population was tracked on the basis of their unique barcodes. We observed dramatic clonal loss in the 1B tetr and iRep1 systems, with the majority lost in the first week following DOX induction (Fig. 2B). The number of clones continued to decrease in DOX+ conditions, with less than 10% surviving 30 days of reprogramming. This was accompanied by a notable increase in cell death (fig. S7A). The removal of DOX at day 14, which forces silencing of exogenous OKMS factor expression, was expected to drive another wave of clonal selection (*18*) and also led to clonal loss (Fig. 2B). Analogous barcoding experiments conducted with the 9DT and primary reprogramming systems produced consistent trends of clonal selection and dominance (see supplementary text and figs. S8 and S9). Notable clonal selection in noninduced MEF cultures was not observed, and there was no corresponding reduction in cell number during reprogramming, indicating that rapid clonal loss was a result of competition in reprogramming (fig. S7, B to D).

Combining cellular barcoding with our published surface-marker profiling strategy (*6*), we measured the fraction of clones that contribute to the MEF, pre-iPSC, and iPSC fractions (see supplementary text and fig. S10). Although the number of clones in an iPSC state increased over time during population reprogramming, not all clones contributed to the iPSC fraction and even fewer survived DOX removal (figs. S7E and S10D). By contrast, our earlier isolated clonal reprogramming study suggested that all clones have iPSC-producing capabilities (Fig. 1B). This discrepancy in clonal iPSC potential in isolated versus mixed conditions signaled that clonal interactions are involved in population dynamics during reprogramming.

We next examined whether clonal loss is accompanied by the survival of a dominant clonal subset. After 30 days of DOX induction, two clones exhibited a clear advantage over their competitors, occupying more than 50% of the culture (Fig. 2C). The primary reprogramming and secondary 9DT systems also gave rise to a few dominant clones, emerging as early as day 14 (figs. S8C and S9C). The surprising observation that a few clones overtook the majority of the cell pool provides a possible basis for the homogeneity seen in mixed clonal populations (Figs. 1B and 2D and fig. S7F). These dominant clones can attain an iPSC state (fig. S10E) and in most, but not all, cases maintain their dominance after DOX removal (Fig. 2C and figs. S7G, S8C, and S9C).

Mathematical model of clonal reprogramming challenges the clonal equipotency hypothesis

It remained unclear whether clonal loss and dominance were due to stochastic reprogramming latencies of MEFs or selective advantages of a subset of clones. To address this question, we

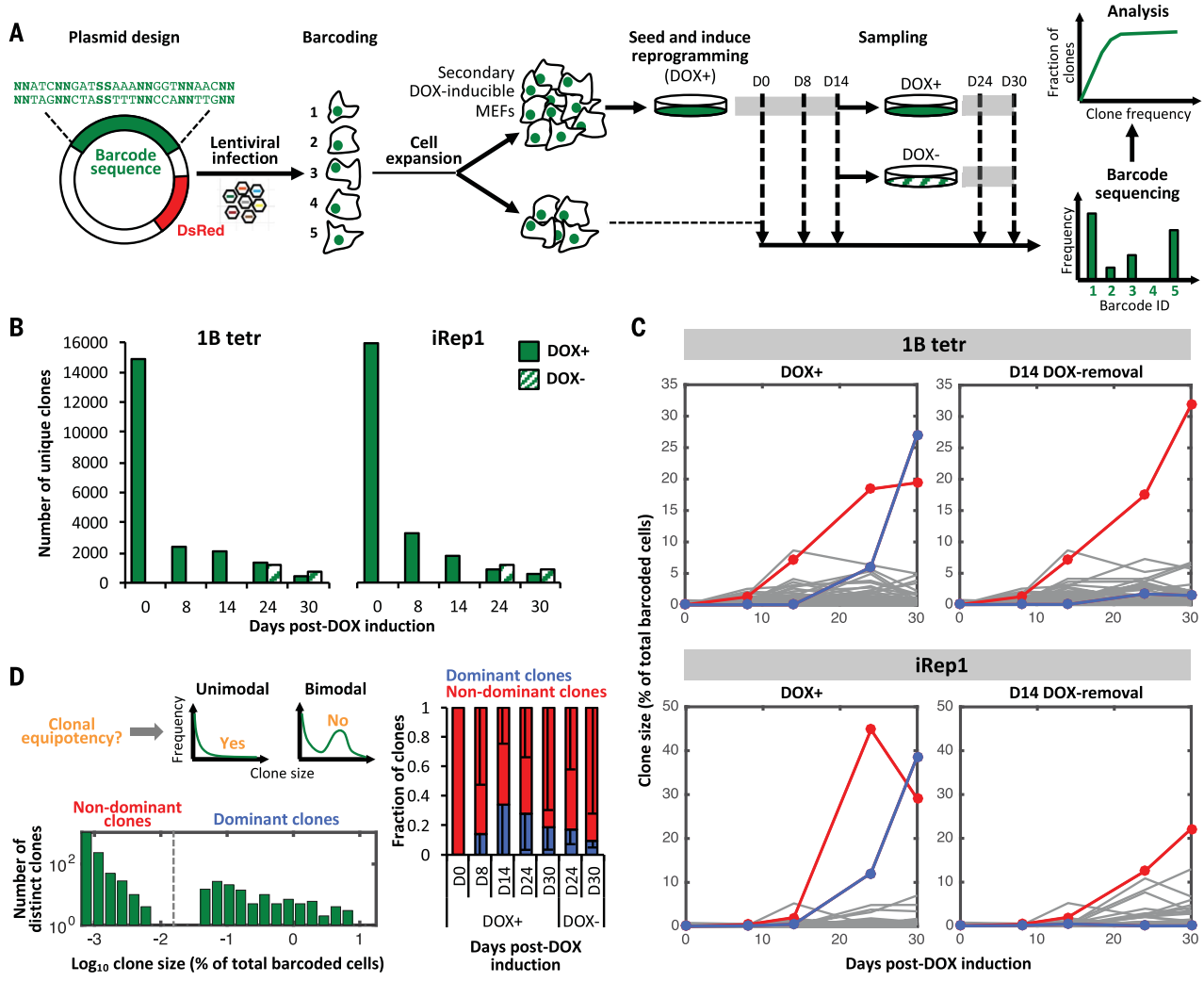


Fig. 2. Cellular barcoding reveals clonal selection in reprogramming. (**A**) The barcoding pipeline, in which lentivirus containing unique DNA barcodes are generated and used to infect secondary MEFs from the 1B tetr and iRep1 reprogramming systems [multiplicity of infection (MOI) ~ 0.05]. Cells are subsequently reprogrammed and regularly sampled (20%) for barcode sequencing to track the temporal behavior of clones throughout reprogramming. DsRed is a red fluorescent protein. (**B**) Number of unique clones (barcodes) over time following DOX induction in the 1B tetr and iRep1 systems. Representative data are from $n = 4$ biological replicates. (**C**) Frequency of individual clones over time following DOX induction in the 1B tetr and iRep1 systems. Clones showing extreme dominance are highlighted in red and blue and indicate the same clone in DOX+ and DOX− cultures. Representative data are from $n = 4$ biological replicates. (**D**) Fraction of clones exhibiting relative dominance (versus nonzero, nondominant clones) over time following DOX induction in the 1B tetr and iRep1 systems. Representative histogram of clone sizes is from the 1B tetr reprogramming system at D24 of DOX+ culture. Bar graph data represent means ± standard deviation of $n = 4$ biological replicates (from 1B tetr and iRep1 systems).

analyzed clone size distributions (*19*). The clonal equipotency model of reprogramming predicts neutral clonal dynamics, whereby dominant clones stochastically initiate reprogramming earlier and therefore have more time to contribute to the reprogrammed population. We observed a bimodal distribution of dominant versus nondominant clones (Fig. 2D and figs. S7H, S8D, and S9D), which is incongruous with the assumption of clonal equipotency because neutral dynamics are expected to produce a unimodal clone size distribution (fig. S1A) (*8*), leading us to hypothesize that clonal dominance is not exclusively driven by stochastic neutral dynamics. To test this, we developed a stochastic model of cellular reprogramming with clonal resolution (fig. S11) and compared the predictions of our model with experimental observations (see methods and figs. S11 to S13).

In silico simulations of clone size distributions in stochastic reprogramming diverged significantly from experimental observations by day 8 (fig. S11, C and D). Notably, the in silico model gave rise to a unimodal distribution of clone sizes, as expected. Given that the most dramatic changes in clone size distributions in vitro occurred within the first week of reprogramming (fig. S11, C and D), accompanied by the greatest clonal loss (Fig. 2B), we next tested whether the neutral model could capture clonal dynamics following the first week of reprogramming. We ran the model with initial conditions from day 8 clone distributions and found that the neutral model did not reproduce the experimental observations following the first week of reprogramming (fig. S11E). These findings support the selective nature of clonal dynamics and led us to examine whether the experimentally observed dynamics could be driven by underlying heterogeneity in clonal fitness.

Clonal eliteness arises in the first week of reprogramming

To determine when larger clones emerge and acquire dominance, we further probed the experimental results of our initial barcoding study (Fig. 2) and compared outcomes based on clone size. We sorted clones into four bins of increasing size while keeping the clone size range (in logarithmic scale) consistent between bins (Fig. 3A). This sorting was performed at different time

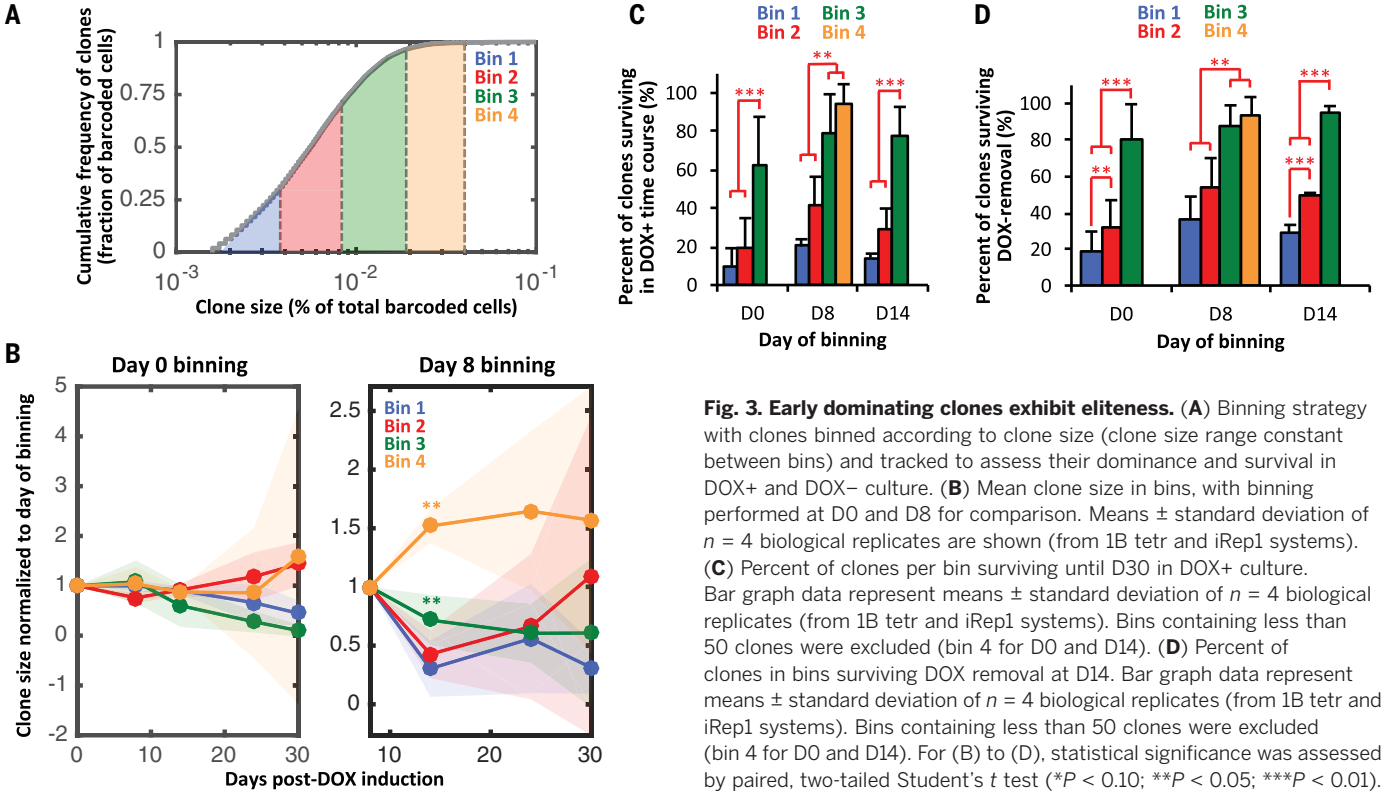


Fig. 3. Early dominating clones exhibit eliteness. (**A**) Binning strategy with clones binned according to clone size (clone size range constant between bins) and tracked to assess their dominance and survival in DOX+ and DOX− culture. (**B**) Mean clone size in bins, with binning performed at D0 and D8 for comparison. Means ± standard deviation of $n = 4$ biological replicates are shown (from 1B tetr and iRep1 systems). (**C**) Percent of clones per bin surviving until D30 in DOX+ culture. Bar graph data represent means ± standard deviation of $n = 4$ biological replicates (from 1B tetr and iRep1 systems). Bins containing less than 50 clones were excluded (bin 4 for D0 and D14). (**D**) Percent of clones in bins surviving DOX removal at D14. Bar graph data represent means ± standard deviation of $n = 4$ biological replicates (from 1B tetr and iRep1 systems). Bins containing less than 50 clones were excluded (bin 4 for D0 and D14). For (B) to (D), statistical significance was assessed by paired, two-tailed Student's t test (*$P < 0.10$; **$P < 0.05$; ***$P < 0.01$).

points, and clone sizes were tracked in each bin over time and then normalized to initial clone size at the time of binning. We tested the propensity for bins with larger clones to become dominant later in culture. The largest clones (bin 4) at days 8 and 14 went on to become relatively dominant compared with the clones in the other three bins (Fig. 3B and fig. S14A). By contrast, clones that were larger at day 0 (owing to stochastic differences in lentiviral barcoding) did not become dominant.

We found that early dominating clones (in bin 4 at day 8) were able to contribute to all four bins throughout reprogramming (fig. S14B). Nevertheless, the majority of late-stage dominant clones (in bin 4 at day 30) were also dominant at day 8. Simulated clone size fluctuations based on the neutral stochastic model also predicted that early dominating clones maintain their dominance (fig. S14B), though the neutral model was unable to capture their dynamics (fig. S11E). Furthermore, early dominating clones were more likely to survive competition over time in DOX+ culture (Fig. 3C). Similar observations were made in experiments grouping clones into four bins of increasing clone size while keeping the number of clones constant between bins (fig. S15, A to E). Applying our binning strategy to the data from the primary reprogramming and secondary 9DT systems, we found similar trends (fig. S16, A and B). Larger clones (in bins 3 or 4) also exhibited an enhanced propensity to contribute to transgene-independent iPSCs in both DOX+ (figs. S14C and S15F) and DOX− culture (Fig. 3D and figs. S15G and S16C), further highlighting their enhanced fitness. Thus, although not all early dominating clones (bin 4 at day 8) dominate in late-stage reprogramming, they have an enhanced probability of doing so.

A priori propensity for clonal eliteness is present in the initial MEF pool

In light of the heterogeneous nature of MEFs (*20*), we examined the possibility that a subset of cells exhibits an a priori eliteness that drives the emergence of dominant clones by day 8 of reprogramming. Following an average of three cell divisions, barcoded MEFs were split into three equally distributed pools: One pool was reserved for day 0 barcode sequencing, whereas the remaining pools were seeded into two flasks and subjected to identical DOX treatment (Fig. 4A and figs. S8A and S9A). Following induction of reprogramming, we tracked clonal loss and confirmed similar selection dynamics in the parallel flasks (fig. S17A).

To determine whether clonal dominance is inherited from the MEF state, we first assessed the correlation in clone sizes between parallel flasks containing MEFs from a common parental origin. In scatter plots (Fig. 4B), the correlation of clones (identified by unique barcodes) between parallel flasks was high (Fig. 4C and figs. S8, E to G; S9, E to G; and S17B)—consistently higher than that expected by chance (fig. S17, D and E).

We applied our binning strategy to the splitting experiment to determine whether a subpopulation of clones was driving the correlation between flasks. We reasoned that if a priori poised clones existed, they would be prominent in the early dominating clones. Our binning strategy revealed that although smaller clones displayed barcode heterogeneity and thus poor correlation, the largest clones were relatively well correlated (fig. S18A). Indeed, removal of bin 4 clones substantially reduced correlation between parallel flasks (fig. S18B). Furthermore, bin 4 clones gave rise to more dominant clones later in culture with comparable outcomes in parallel flasks (figs. S18C and S19A). Similar correlations were observed in primary and secondary 9DT systems (figs. S16, D and E, and S19B). Thus, early dominating clones arise from a subpopulation of poised MEFs, inheriting their potential for dominance and driving the correlations observed in our parallel reprogramming flasks, while also exhibiting reproducible dominance (see supplementary text and fig. S20).

Taken together, our results suggest that a priori eliteness associated with parental MEF status determines a clone's ability to initiate reprogramming quickly and dominate, confirming that the heterogeneous MEF population displays cell-specific reprogramming rates (*21*). Furthermore, the rigorous analysis of clonal selection and dominance dynamics presented here may also inform iPSC production strategies (see supplementary text and fig. S21).

MEFs derived from Wnt1-expressing cells are poised to dominate the reprogramming pool

Our barcoding analysis revealed the existence of a subpopulation of elite MEFs. To uncover the

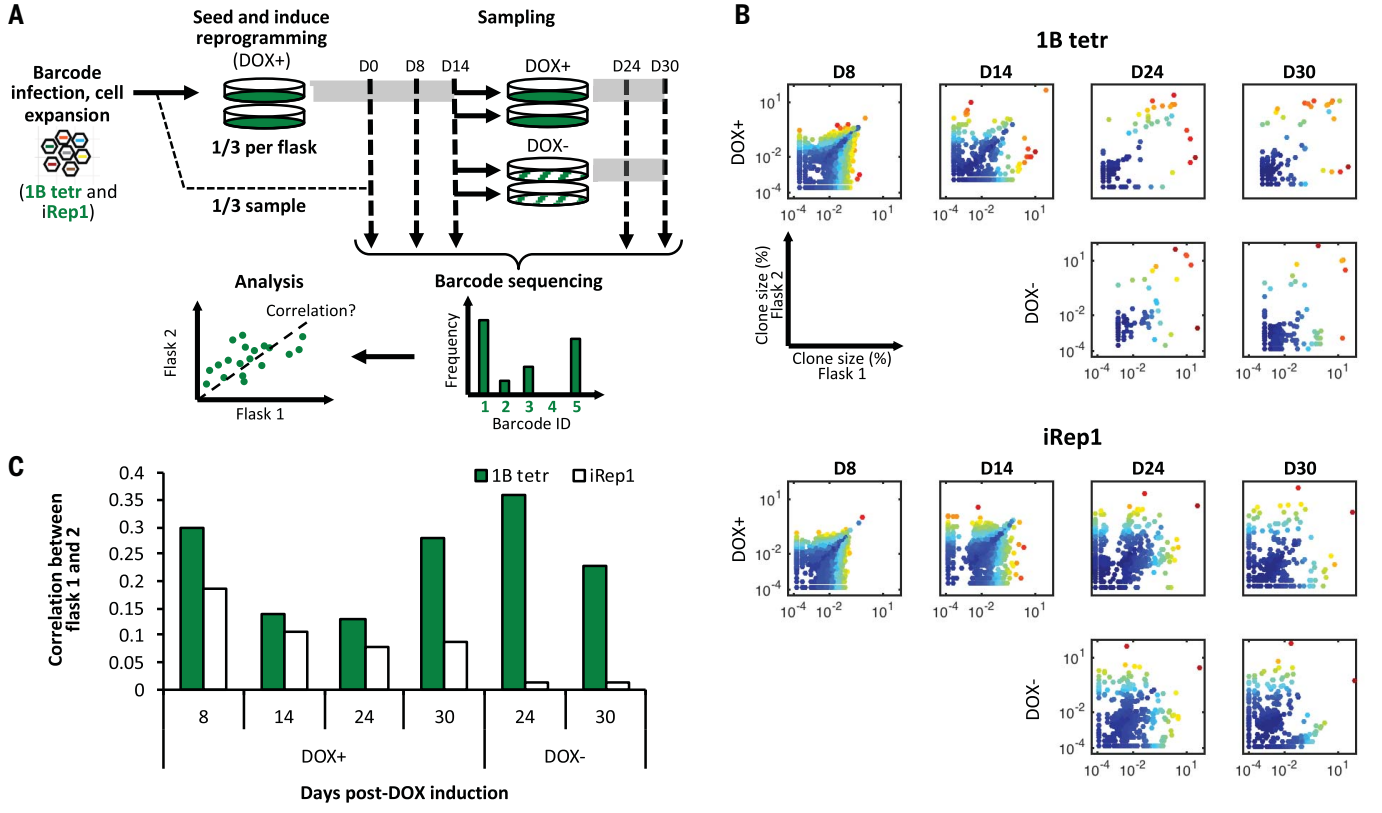


Fig. 4. A subpopulation of a priori elite MEFs gives rise to early dominating clones. (**A**) Population splitting experiment in which barcoded secondary MEFs (expanded for 3 days) from the 1B tetr and iRep1 reprogramming systems are split at D0 of DOX induction into parallel flasks and sampled regularly for barcode sequencing. Parallel flasks were assessed for similarity in outcomes of common clones. (**B**) Scatter plots of individual clone sizes (percentage of total barcoded cells) in parallel reprogramming flasks over time following DOX induction in the 1B tetr and iRep1 systems. Color shows the distance of each point to the diagonal (blue is shorter; red is longer). (**C**) Correlation between clone sizes in parallel flasks over time following DOX induction in the 1B tetr and iRep1 systems.

identity of these MEFs, we tested whether the neural crest (NC) population in skin has an a priori ability to dominate reprogramming (see supplementary text).

We developed a lineage tracing strategy to retrospectively identify MEFs based on Wnt1 expression, which has been suggested to mark NC-derived cells (*22*, *23*) (see methods and fig. S22). Upon isolation and serial passaging, the fraction of NC-derived MEFs became increasingly prominent in nonreprogramming culture, suggesting their competitive advantage (fig. S22B). We next infected the population containing both NC and non-NC–derived MEFs with lentivirus carrying the OKMS factors and scored the origin of emerging SSEA1+ clones. After 3 weeks of reprogramming, 100% of reprogramming colonies originated from NC-derived MEFs (fig. S22C). Furthermore, upon cell sorting, we found a higher efficiency of reprogramming induction in NC-derived clones (fig. S22D).

To validate these preliminary findings from low-efficiency primary reprogramming experiments, we next explored the dynamics of NC-derived MEFs in an efficient secondary reprogramming system (Fig. 5A and methods). The prominence of NC-derived [yellow fluorescent protein positive (YFP+)] MEFs increased in reprogramming culture, where the YFP+ fraction overtook the population (Fig. 5B and fig. S22E). This suggested that NC-derived cells exhibited a competitive advantage in both MEF and reprogramming culture and prompted us to further assess the dynamics of their dominance. We conducted mixing studies, controlling the initial fraction of YFP+ to YFP− cells while keeping the total cell number constant (Fig. 5C). The fraction of YFP+ cells increased over the course of reprogramming and overtook the population in almost all cases, regardless of seeding frequency. Given the propensity of YFP+ cells to overtake reprogramming populations, we next assessed their reprogramming potential in isolated clonal studies and found that NC-derived MEFs exhibited a significantly higher ability to initiate reprogramming (Fig. 5D).

Given the results of our isolated clonal experiments, we next asked whether reprogramming success was dependent on NC-derived MEF frequency at day 0 of reprogramming. As expected, higher initial levels of YFP+ MEFs gave rise to higher levels of SSEA1+ cells at earlier time points (Fig. 5E). Whereas the fractions of SSEA1+ cells arising from the YFP+ and YFP− subpopulations were comparable in these studies (fig. S22F), the overall SSEA1+ pool was predominantly composed of YFP+ cells (fig. S22G). This suggests that although YFP− MEFs can give rise to reprogramming cells, the YFP+ cells overtake the reprogramming compartment. The eliteness of YFP+ reprogramming clones may be linked to a higher reprogramming propensity as well as a higher baseline competitive advantage, as seen from the DOX− controls (fig. S22B). The observation that YFP− cells can reprogram in isolation but have limited contribution to population-based reprogramming underscores the effects of competition on reprogramming dynamics.

As a final validation, we expanded our stochastic model of reprogramming in which all clones have the same potential to reprogram (fig. S11), which did not accurately capture experimental dynamics, to a two-population stochastic model. In our new model, two MEF subpopulations are defined (A and B), where B has an a priori propensity to dominate the reprogramming pool, as in NC-derived MEFs (Fig. 5F). This two-population reprogramming model accurately captured clonal dynamics (Fig. 5G), representing a drastic improvement over our original model. The model quantified the relative eliteness of NC-derived clones and demonstrated that these cells are consistently more competitive than their nonelite counterparts (Fig. 5H). We also found no notable improvement

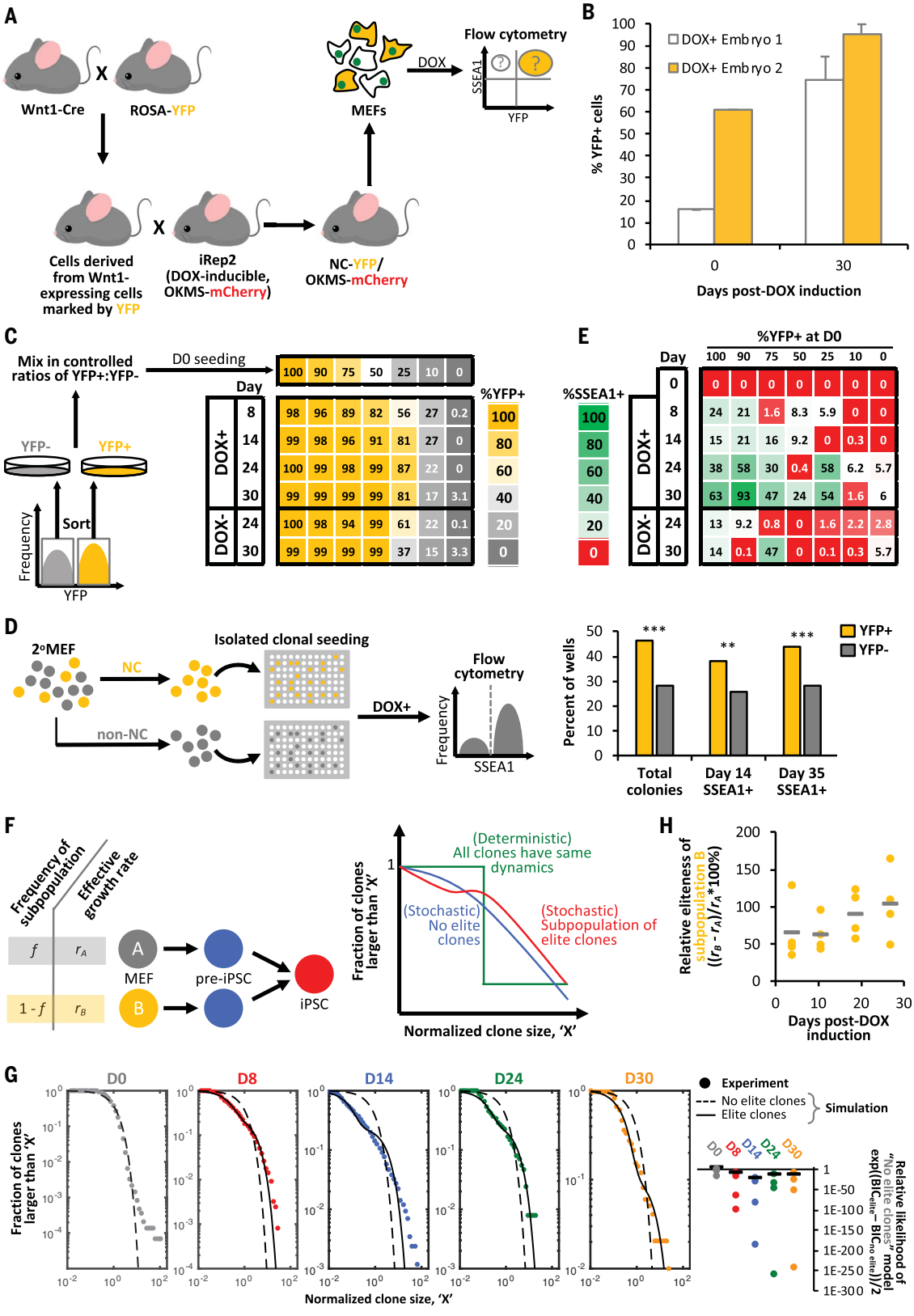


Fig. 5. NC-derived MEFs dominate the reprogramming population. (**A**) Secondary reprogramming system labeling NC-derived MEFs. (**B**) Fraction of NC cells after 30 days of reprogramming. Means ± standard deviation of $n = 3$ biological replicates are shown. (**C**) Controlled mixing studies of NC and non-NC MEFs. The heatmap shows the fraction of NC (YFP+) cells over time. Representative results are from $n = 2$ biological replicates. (**D**) Reprogramming efficiency of NC and non-NC MEFs, assayed by isolated single-cell reprogramming assay. Statistical significance was assessed by Z test (**$P < 0.01$; ***$P < 0.001$). (**E**) Reprogramming fractions in controlled mixing studies, assayed by SSEA1 expression. Representative results are from $n = 2$ biological replicates. (**F**) Compartmental probabilistic model of reprogramming, including an elite MEF fraction (subpopulation B, corresponding to NC-derived MEFs). Expected clone size distribution plots for the new (elite subpopulation) and old (no elite subpopulation) models are shown. (**G**) Experimental clone size distributions from the 1B tetr system, with model-generated curves overlaid. Bayesian information criterion (BIC) quantifies improved fit of the two-population model. Means ± standard deviation of $n = 4$ biological replicates are shown (from 1B tetr and iRep1 systems). (**H**) Relative eliteness of subpopulation B. Data points represent $n = 4$ biological replicates; the gray bar shows the mean (from 1B tetr and iRep1 systems). r_A and r_B, effective growth rates of subpopulations A and B.

in the model's ability to accurately recapitulate experimental data by adding more than two MEF subpopulations (fig. S23A), supporting the observation that the single identified NC-derived MEF fraction is the primary source for elite clones. Furthermore, when applied to our day 0 population splitting experiment, the two-population model recapitulated the range of observed positive correlations between parallel flasks (fig. S23B) as well as the observation that larger clones make the main contribution to the correlation (fig. S23C), whereas the neutral model only produced correlations of zero. Fitting the experimental correlations to the two-population model also allowed us to estimate the frequency of elite MEFs (fig. S23D), which was bound by the measured fraction of NC-derived MEFs. We also noted that the model was able to accurately capture the dynamics of clonal loss we observed experimentally (fig. S23E). We next used the two-population model to understand if there are two phases to clonal dynamics—an early phase governed by clonal selection and a later phase resulting from neutral or stochastic, clonal loss. Indeed, our earlier observation that clone size distributions shift most dramatically between days 0 and 8 (fig. S11D) suggested this possibility. As expected, the model predicted the selective advantage of the elite reprogramming subpopulation to be larger than the nonelite fraction (fig. S23F). Furthermore, the time-varying nature of the effective growth rates of the two subpopulations (see methods) suggested that the entire reprogramming process is governed by selective dynamics, consistent with our earlier finding that the later phase of reprogramming could not be captured by neutral clonal dynamics (fig. S11E).

These observations confirmed that individual MEFs exhibit unequal propensities for reprogramming induction. This led us to uncover a mechanism for the observation of a priori reprogramming eliteness, which is driven by a subpopulation of MEFs derived from Wnt1-expressing cells, likely corresponding to an NC population. Additionally, an exploration of growth and death dynamics in isolated versus coculture of NC and non-NC cells provided some evidence for direct interactions between these competing populations (see supplementary text and fig. S24).

Discussion

We aimed to uncover the connection between intrinsic cell identity, the genetically encoded potential of a cell, and cell-extrinsic community effects (Fig. 6). During reprogramming, whether inequalities in cell fitness drive non-neutral clonal drift in interacting cells remains unclear. We found a poised subpopulation of somatic cells that gives rise to dominant clones within the first week of population reprogramming. These early dominating clones are fitter, exhibiting a higher probability of dominance in later culture and an increased propensity to reprogram. Indeed, our lineage tracing experiments uncovered the identity of a previously unknown subpopulation of MEFs from the Wnt1+ (neural crest) lineage that preferentially gives rise to elite reprogramming clones. These observations align with previous reports of the heterogeneous reprogramming potential of MEFs (*21*, *24*, *25*). Our data suggest that the eliteness of NC-derived MEFs may be linked to their ability to initiate reprogramming more efficiently, allowing them to transition to an iPSC state more rapidly and escape the MEF and pre-iPSC states, which experience increased cell death.

Our findings make clear that competitive interactions between reprogramming clones drive population dynamics in a manner that is different from isolated clonal studies, suggesting that the clonal equipotency paradigm—in which somatic cells have an equal potential to attain an iPSC state in isolation—may not govern clonal dynamics in the population setting. These

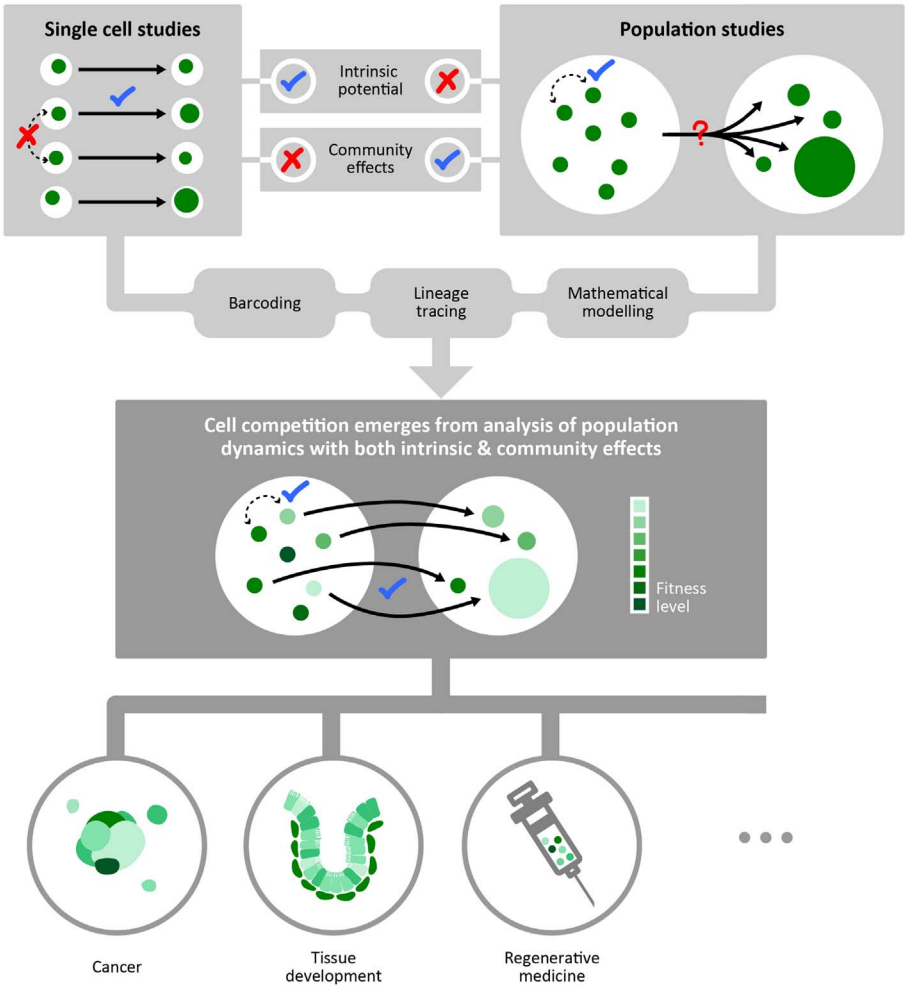


Fig. 6. Competition in reprogramming emerges as a relevant parameter that bridges our understanding of population dynamics and single-cell potential. By exploring the dynamics of clones within multicellular populations, cell competition emerges as a mechanism by which cells with hidden, context-specific eliteness emerge to occupy and dominate the reprogramming niche. Cell competition serves to connect our understanding of the genetically encoded potential of a cell, quantified by single-cell studies, and the behavior of these cells within the community context. This paradigm can help to inform our understanding of developmental and disease systems as well as strategies for controlling the competitive potential of cell therapies.

interactions may be influenced by indirect competition for limited culture nutrients and space or by direct competition between cells. Indeed, direct cell competition has been reported to drive population homogeneity in PSCs in vivo (*11*) and in vitro (*13*, *14*), where Myc^{high} cells directly induce apoptosis in their Myc^{low} neighbors. Given that Myc overexpression is involved in the induction of reprogramming, Myc-mediated competition between reprogramming cells is not unexpected. Indeed, population-level Myc expression reportedly increases over the reprogramming process in the presence of exogenous OKMS factor expression, suggesting that these cells may overtake the culture (*5*). To gain clarity on the role of Myc in this process, we need a deeper understanding of the molecular mechanisms of fitness in reprogramming, which requires as yet undeveloped tools for the isolation and single-cell sequencing of rare populations of specific barcoded cells. This technical challenge currently poses a barrier to the omics characterization of elite reprogramming clones.

We found that a few clones overtake the reprogramming population. Although it remains unclear whether these clones inherit their eliteness potential from the MEF state, it is possible that they acquire their dominance in a culture-induced manner. Indeed, the predictable behavior of the clones in parallel reprogramming flasks could be due to the acquisition of heritable changes that confer enhanced cellular fitness. Previous reports have pointed to genomic aberrations in PSCs acquired in culture (*26*), and it has been reported that passaging results in the loss of most of these mutations (*27*). Indeed, active competition in PSC culture has been shown to remove genomic aberrations (*13*). These observations underline that emergence of culture-induced eliteness is a potential secondary source of competition to consider in long-term culture.

Our data also support the possibility that NC stem cells present in the MEF population are the source of these top dominating clones. NC stem cells are developmentally related to neural stem cells, which have been shown to reprogram to an iPSC state through the overexpression of Oct4 alone (*28*). Notably, NC stem cells can be expanded long-term and express core pluripotency factors, which may poise them to overtake the population. Indeed, our data may support at least three time-dependent selection events, all of which may involve NC cells: (i) NC MEFs dominate somatic culture before reprogramming, (ii) NC clones are further selected through competition with non-NC cells during early reprogramming, and (iii) reprogrammed NC stem cells and their progeny become dominant and overtake later cultures. Overall, it is likely that iPSCs are not equivalent in terms of fitness, thus meriting the careful selection of PSC clones for regenerative medicine applications (*29*).

This study uncovers cell competition as a mechanism for cells with context-specific eliteness to occupy the reprogramming niche. These findings support a paradigm shift, turning competition into a generalizable parameter for understanding multicellular developmental and disease systems (Fig. 6). To this end, reprogramming may serve as a model to probe the genetic and epigenetic basis of cell fitness, providing key insights toward controlling competition for regenerative medicine (*29*). This will be critical for synthetic biology strategies that aim to engineer cells with predictable and desirable behaviors in multicellular contexts (*30*).

Methods summary

For reprogramming, DOX-inducible secondary MEF 1B cells [isolated as previously described (*15*) from tetraploid complementation], MEF iRep1 cells [derived by previous methods (*15*)], MEF 9DT cells [similar to that previously reported (*17*)], or primary MEFs were used. Following cellular barcoding by viral infection using an adapted version of a previously reported cellular barcoding strategy (*16*), MEFs were subsequently passaged once to allow for an average of ~threefold expansion and subsequently split, whereby cells were sampled for initial barcode profiling and the remaining cells were seeded for reprogramming induction. Following DOX induction, cultures were sampled regularly for sequencing. Cells were transitioned to DOX-free culture to assess clone contributions to the generation of transgene-independent iPSCs.

To assess barcode survival and abundance, we conducted targeted DNA paired-end sequencing followed by analysis with an in-house bioinformatics analysis pipeline using custom Perl and MATLAB scripts.

For NC lineage studies, we created a mouse strain containing the transgene Wnt1-Cre and floxed ROSA-tdTomato. Upon Cre expression, Wnt1-expressing cells and their progeny become irreversibly labeled. To develop a more efficient reprogramming system with NC tracing, a homozygous ROSA-YFP mouse was crossed to a heterozygous Wnt1-Cre. The Wnt1-Cre/ROSA-YFP mice were then crossed to the highly efficient secondary reprogramming system iRep2 [derived by previous methods (*15*)] mouse line.

Mathematical modeling of clonal dynamics was conducted with custom Python and MATLAB scripts. Calculations for statistical significance were performed using Excel and MATLAB software.

REFERENCES AND NOTES

1. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006). doi: 10.1016/j.cell.2006.07.024; pmid: 16904174
2. P. Samavarchi-Tehrani *et al.*, Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010). doi: 10.1016/j.stem.2010.04.015; pmid: 20621051
3. J. M. Polo *et al.*, A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012). doi: 10.1016/j.cell.2012.11.039; pmid: 23260147
4. J. Hansson *et al.*, Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Reports* **2**, 1579–1592 (2012). doi: 10.1016/j.celrep.2012.10.014; pmid: 23260666
5. S. M. I. Hussein *et al.*, Genome-wide characterization of the routes to pluripotency. *Nature* **516**, 198–206 (2014). doi: 10.1038/nature14046; pmid: 25503233
6. N. Shakiba *et al.*, CD24 tracks divergent pluripotent states in mouse and human cells. *Nat. Commun.* **6**, 7329 (2015). doi: 10.1038/ncomms8329; pmid: 26076835
7. J. O'Malley *et al.*, High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* **499**, 88–91 (2013). pmid: 23728301
8. J. Hanna *et al.*, Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595–601 (2009). doi: 10.1038/nature08592; pmid: 19898493
9. Y. Buganim *et al.*, Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012). doi: 10.1016/j.cell.2012.08.023; pmid: 22980981
10. G. Morata, P. Ripoll, Minutes: Mutants of Drosophila autonomously affecting cell division rate. *Dev. Biol.* **42**, 211–221 (1975). doi: 10.1016/0012-1606(75)90330-9; pmid: 1116643
11. C. Clavería, G. Giovinazzo, R. Sierra, M. Torres, Myc-driven endogenous cell competition in the early mammalian embryo. *Nature* **500**, 39–44 (2013). doi: 10.1038/nature12389; pmid: 23842495
12. M. Amoyel, E. A. Bach, Cell competition: How to eliminate your neighbours. *Development* **141**, 988–1000 (2014). doi: 10.1242/dev.079129; pmid: 24550108
13. M. Sancho *et al.*, Competitive interactions eliminate unfit embryonic stem cells at the onset of differentiation. *Dev. Cell* **26**, 19–30 (2013). doi: 10.1016/j.devcel.2013.06.012; pmid: 23867226
14. C. Díaz-Díaz *et al.*, Pluripotency surveillance by Myc-driven competitive elimination of differentiating cells. *Dev. Cell* **42**, 585–599.e4 (2017). doi: 10.1016/j.devcel.2017.08.011; pmid: 28919206
15. K. Woltjen *et al.*, *piggyBac* transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009). doi: 10.1038/nature07863; pmid: 19252478
16. A. Gerrits *et al.*, Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010). doi: 10.1182/blood-2009-06-229757; pmid: 20093403
17. H. Han *et al.*, MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* **498**, 241–245 (2013). doi: 10.1038/nature12270; pmid: 23739326
18. A. Golipour *et al.*, A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* **11**, 769–782 (2012). doi: 10.1016/j.stem.2012.11.008; pmid: 23217423
19. S. Goyal, S. Kim, I. S. Y. Chen, T. Chou, Mechanisms of blood homeostasis: Lineage tracking and a neutral model of cell populations in rhesus macaques. *BMC Biol.* **13**, 85 (2015). doi: 10.1186/s12915-015-0191-8; pmid: 26486451
20. P. K. Singhal *et al.*, Mouse embryonic fibroblasts exhibit extensive developmental and phenotypic diversity. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 122–127 (2016). doi: 10.1073/pnas.1522401112; pmid: 26699463
21. L. L. Liu *et al.*, Probabilistic modeling of reprogramming to induced pluripotent stem cells. *Cell Reports* **17**, 3395–3406 (2016). doi: 10.1016/j.celrep.2016.11.080; pmid: 28009305
22. Y. Chai *et al.*, Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. *Development* **127**, 1671–1679 (2000). pmid: 10725243
23. P. S. Danielian, D. Muccino, D. H. Rowitch, S. K. Michael, A. P. McMahon, Modification of gene activity in mouse embryos *in utero* by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8**, 1323–1326 (1998). doi: 10.1016/S0960-9822(07)00562-3; pmid: 9843687
24. S. Guo *et al.*, Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156**, 649–662 (2014). doi: 10.1016/j.cell.2014.01.020; pmid: 24486105
25. M. Pour *et al.*, Epigenetic predisposition to reprogramming fates in somatic cells. *EMBO Rep.* **16**, 370–378 (2015). doi: 10.15252/embr.201439264; pmid: 25600117
26. U. Ben-David, N. Benvenisty, Y. Mayshar, Genetic instability in human induced pluripotent stem cells: Classification of causes and possible safeguards. *Cell Cycle* **9**, 4603–4604 (2010). doi: 10.4161/cc.9.23.14094; pmid: 21099357
27. S. M. Hussein *et al.*, Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011). doi: 10.1038/nature09871; pmid: 21368824
28. J. B. Kim *et al.*, Oct4-induced pluripotency in adult neural stem cells. *Cell* **136**, 411–419 (2009). doi: 10.1016/j.cell.2009.01.023; pmid: 19203577
29. N. Shakiba, P. W. Zandstra, Engineering cell fitness: Lessons for regenerative medicine. *Curr. Opin. Biotechnol.* **47**, 7–15 (2017). doi: 10.1016/j.copbio.2017.05.005; pmid: 28551499

30. B. P. Teague, P. Guye, R. Weiss, Synthetic morphogenesis. *Cold Spring Harb. Perspect. Biol.* **8**, a023929 (2016). doi: 10.1101/cshperspect.a023929; pmid: 27270296

ACKNOWLEDGMENTS

The authors thank the Bystrykh laboratory for sharing their barcoding plasmids. We also thank C. L. Bauwens for her editing, K. Chan for assisting with sample sequencing, M. Parsons for assisting with flow cytometry, T. Yin for performing cell culture and sample preparation for flow cytometry, and J. Ma for artistic contributions to the figures. We acknowledge the assistance and support of laboratory colleagues and collaborators. **Funding:** This work has been supported by Medicine by Design (University of Toronto), CFREF program, CIHR (MOP 57885), and CIHR Foundation grants to P.W.Z., D.v.d.K., A.N., and J.L.W. N.S. is the recipient of the NSERC Vanier Canada Graduate Scholarship. A.N. is the Tier 1 Canada Research Chair in Stem Cells and Regeneration. P.W.Z. is the Canada Research Chair in Stem Cell Bioengineering. **Author contributions:** N.S., A.F., G.J., J.L.W., D.v.d.K., and P.W.Z. conceived and designed the experiments. N.S. and P.W.Z. wrote the manuscript. A.N. contributed to study design. N.S. and G.J. performed all barcoding experiments, cell culture, sequencing, and bioinformatics work. A.F. developed the neural crest lineage tracing system. S.M. and S.G. conducted in silico simulations and mathematical modeling studies. L.D. developed the lentiviral barcoding system. J.E. and M.C.P. created the 1B tetr, iRep1, and iRep2 secondary reprogramming systems. G.J. and D.T. designed and characterized the primary and secondary 9DT reprogramming systems. **Competing interests:** The authors declare that they have no competing interests.

Data and materials availability: All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6438/eaan0925/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S24
Table S1
References (*31*–*39*)

3 March 2017; resubmitted 2 August 2018
Accepted 25 February 2019
Published online 21 March 2019
10.1126/science.aan0925

Science

Cell competition during reprogramming gives rise to dominant clones

Nika Shakiba, Ahmed Fahmy, Gowtham Jayakumaran, Sophie McGibbon, Laurent David, Daniel Trcka, Judith Elbaz, Mira C. Puri, Andras Nagy, Derek van der Kooy, Sidhartha Goyal, Jeffrey L. Wrana and Peter W. Zandstra

Science **364** (6438), eaan0925.
DOI: 10.1126/science.aan0925originally published online March 21, 2019

Domination in the stem cell world

A Nobel Prize–winning discovery showed that specialized cells can be genetically reprogrammed into stem cells, thus gaining the ability to become any cell type in the body. But what happens during reprogramming is not completely understood. Shakiba *et al.* used experimental and mathematical approaches to show that skin cells compete during reprogramming, eliminating one another as the population progresses toward the stem cell state (see the Perspective by Wolff and Purvis). The "winners" are a special class of skin cells originating from the neural crest. Cells of this type normally emerge during embryonic development and migrate into various tissues, including the skin, muscle, and nervous system.

Science, this issue p. eaan0925; see also p. 330