npg

ORIGINAL ARTICLE

Evidence for a critical role of gene occlusion in cell fate restriction

Jedidiah Gaetz[1], Kayla L Clift[1], Croydon J Fernandes[1], Frank Fuxiang Mao[1, 2], Jae Hyun Lee[1], Li Zhang[1], Samuel W Baker[1], Timothy J Looney[1], Kara M Foshay[1], Wei-Hua Yu[2], Andy Peng Xiang[2], Bruce T Lahn[1]

[1]*Department of Human Genetics, University of Chicago, Howard Hughes Medical Institute, Chicago, IL 60637, USA; [2]Center for Stem Cell Biology and Tissue Engineering, Sun Yat-sen University, Guangzhou, Guangdong 510080, China*

The progressive restriction of cell fate during lineage differentiation is a poorly understood phenomenon despite its ubiquity in multicellular organisms. We recently used a cell fusion assay to define a mode of epigenetic silencing that we termed "occlusion", wherein affected genes are silenced by *cis*-acting chromatin mechanisms irrespective of whether *trans*-acting transcriptional activators are present. We hypothesized that occlusion of lineage-inappropriate genes could contribute to cell fate restriction. Here, we test this hypothesis by introducing bacterial artificial chromosomes (BACs), which are devoid of chromatin modifications necessary for occlusion, into mouse fibroblasts. We found that BAC transgenes corresponding to occluded endogenous genes are expressed in most cases, whereas BAC transgenes corresponding to silent but non-occluded endogenous genes are not expressed. This indicates that the cellular milieu *in trans* supports the expression of most occluded genes in fibroblasts, and that the silent state of these genes is solely the consequence of occlusion *in cis*. For the BAC corresponding to the occluded myogenic master regulator *Myf5*, expression of the *Myf5* transgene on the BAC triggered fibroblasts to acquire a muscle-like phenotype. These results provide compelling evidence for a critical role of gene occlusion in cell fate restriction.
Keywords: cell fate restriction; occlusion; bacterial artificial chromosome
Cell Research (2012) **22**:848-858. doi:10.1038/cr.2011.190; published online 29 November 2011

Introduction

In multicellular organisms, a single cell – the fertilized egg – undergoes proliferation and differentiation to give rise to all cell types in the body. During this process, the cells not only acquire the phenotypic identities of the lineages to which they have become committed but also lose the potential for all other lineages [1, 2]. The latter phenomenon, known as cell fate restriction, is manifested as the general absence of dedifferentiation or transdifferentiation under physiological conditions. As yet, little is known about what underlies cell fate restriction. While the literature suggests the involvement of both transcription factors acting *in trans* and chromatin components acting *in cis* [3-8], the nature and extent by which either *trans* or *cis* mechanism contributes to cell fate restriction

remain poorly understood. In particular, although many studies have documented extensive chromatin changes during differentiation [6, 8], it is not clear whether these *cis* changes are themselves the primary cause of cell fate restriction, or whether they are merely the downstream effect of *trans* mechanisms. Indeed, there is a dearth of experimental approaches for testing whether gene silencing is due to *cis* or *trans* mechanism in various developmental and physiological contexts.

We recently developed a cell fusion assay to assess whether the primary causal mechanism of gene silencing lies *in trans* or *cis* [9, 10]. Specifically, the assay reveals that the transcriptional potential of a gene can exist in one of two states: competent or occluded. A competent gene is responsive to *trans*-acting factors in the cellular milieu, such that it is active when appropriate transcriptional activators are present, though it can also be silent when activators are absent or repressors are present. In contrast, an occluded gene is blocked by *cis*-acting, chromatin-based mechanisms from responding to *trans*-acting factors, such that it remains silent irrespective of

whether transcriptional activators are present in the cellular milieu. While expressed genes are obviously competent, the cell fusion assay reveals that silent genes can be further divided into those that are still competent and those that are occluded (we will hereon refer to genes that are silent but still competent as activatable genes). Activatable genes are silenced by *trans* mechanisms (i.e., absence of transcriptional activators or presence of repressors), and can potentially be activated again when the cellular milieu changes. In contrast, occluded genes are silenced by *cis* mechanisms and cannot be activated even in the presence of transcriptional activators.

Our definition of "occlusion" bears an important distinction from the commonly used term "epigenetic silencing". In the literature, a gene is often said to be epigenetically silenced when its silent state is associated with certain chromatin marks. However, it is typically not clear whether these chromatin marks are the cause or effect for silencing. In contrast, the term "occlusion" refers only to the type of silencing where *cis*-acting chromatin mechanisms are the primary causal agent of the silent state irrespective of whether transcriptional activators for the gene are present in the cell [9, 10].

The chromatin-based mechanisms causally responsible for the occlusion of genes are not known, but there are candidates, including any of the numerous covalent histone modifications, DNA methylation, binding of repressive complexes, such as Polycomb, subnuclear localization, or other molecular processes and factors that are currently unknown.

We hypothesized that the occlusion of lineage-inappropriate genes might contribute to the restriction of cell fate [9, 10]. Specifically, we proposed that as cells commit to a particular lineage, some of the key genes that specify alternative lineages are occluded, such that these cells lose their ability to express these genes. Implicit to the hypothesis is the assumption that some of the lineage-inappropriate genes subjected to occlusion are present in a cellular milieu that is supportive of their expression, such that they would be expressed if it were not for their occluded state [10].

We decided to test this hypothesis by introducing bacterial artificial chromosomes (BACs) into mouse tail fibroblasts (these cells are designated 129TF). BACs are typically large enough to accommodate all or most *cis*-regulatory regions of a gene, and the lack of chromatin modifications ensures that BAC transgenes are not occluded when first entering the cell. As such, the expression status of the BAC transgenes should provide a readout of the *trans*-acting milieu of the cell. Active expression of a BAC transgene indicates a transcriptionally supportive milieu for that gene (i.e., the cell contains transcriptional activators for the gene), whereas the lack of expression indicates a transcriptionally non-supportive milieu (i.e., the cell lacks activators or contains repressors for the gene). We show here that BAC transgenes corresponding to occluded endogenous genes are frequently expressed, whereas BAC transgenes corresponding to activatable endogenous genes are not expressed. This means that the cellular milieu *in trans* is actually supportive of the expression of most occluded genes (i.e., transcriptional activators for these genes are present in the cell), and the only reason for their silent state is occlusion *in cis*. We further show that when the BAC corresponding to the occluded myogenic master regulator *Myf5* was introduced into fibroblasts, expression of the *Myf5* transgene on the BAC triggered a profound change of fate identity in these cells towards a muscle-like phenotype, indicating that occlusion of endogenous *Myf5* is essential for preventing fibroblasts from activating muscle programs. These findings reveal a critical role of occlusion in safeguarding somatic cell fate.

Results

Identification of occluded and activatable genes for BAC transgene analysis

We have been systematically mapping occluded and activatable genes in 129TF by fusing them to a variety of mouse and rat cell types, including mouse myoblasts (C2C12), rat myoblasts (L6), rat cardiomyoblasts (H9), rat osteosarcoma cells (UMR), and rat chondrocytes (IRC) (manuscript submitted). From the occluded and activatable genes that we identified in these fusions, 10 genes from each category were chosen for the BAC transgene analysis. We gave priority to genes involved in myogenesis because this is a well-studied system and was the focus of our recently published studies on occlusion [9, 11], but we also chose a number of genes randomly.

We first confirmed the occluded or activatable status of these genes in 129TF using the "RT-PCR-Seq" protocol. RT-PCR primers were designed to amplify a gene of interest from both 129TF and its fusion partner, flanking nucleotide sites that differ between the two cell types. RT-PCR product from fused cells was then sequenced, and the allele composition at these nucleotide sites was used to assess whether the gene was expressed from one or both of the two genomes in fused cells. For activatable genes in 129TF, RT-PCR showed a lack of expression in 129TF and robust expression in its fusion partner prior to fusion. After fusion, expression in fused cells came from both the 129TF genome and the fusion partner's genome, indicating that the 129TF copies of these genes were activated upon fusion (Figure 1A). For occluded genes in

129TF, there is a lack of expression in 129TF, but robust expression in its fusion partner prior to fusion, just like activatable genes. However, after fusion, expression in fused cells came only from 129TF's fusion partner and not from the 129TF genome, indicating that the 129TF copies of these genes failed to turn on in fused cells even though their orthologous copies in the fusion partner's genome were expressed (Figure 1B). We also included *Myf6* and *Chrnd* in the BAC transgene analysis. The occlusion status of these two genes in 129TF is unknown. However, in other mouse fibroblast lines that we examined, *Myf6* is consistently occluded while *Chrnd* is activatable ([9]; unpublished data). We therefore assumed that the same is true in 129TF.

In total, 11 occluded and 11 activatable genes were subjected to the BAC transgene analysis. We obtained rat BACs corresponding to these genes and transfected them into 129TF (Supplementary information, Table S1). We used rat (as opposed to mouse) BACs because this allowed us to easily distinguish transgene expression from endogenous gene expression by exploiting sequence divergence between rat and mouse orthologs, while also minimizing potential interspecies genetic incompatibility given the close evolutionary relationship between mouse and rat. The BAC inserts ranged in size from 173 to 252 kb, and for each gene, the corresponding BAC carried the entire transcriptional unit plus a minimum of 40 kb upstream and 27 kb downstream sequences (Supplementary information, Table S1). The great majority of BACs also contained additional genes that flanked the genes of interest. It is therefore likely that most, if not all, of the regulatory sequences in the genes of interest are present on the BACs. For ease of description, BAC transgenes corresponding to activatable endogenous genes will be referred to as "a-transgenes", whereas BAC transgenes corresponding to occluded endogenous genes will be called "o-transgenes."

We modified each BAC by inserting a neomycin-selectable cassette into the vector backbone, transfected it into 129TF cells, and selected for stable BAC integration using neomycin. For each BAC, we chose up to three neomycin-resistant clones and performed RT-PCR with mouse/rat common primers for the genes of interest, followed by sequencing of RT-PCR product to examine whether the expression, if any, came from the rat BAC transgene or from the endogenous mouse gene.

BAC transgenes corresponding to activatable endogenous genes are silent

Because the silent state of the activatable genes is supposedly due to the presence of a transcriptionally non-supportive milieu acting *in trans*, we expected that

when an a-transgene is introduced into the same milieu, it should behave identically as the endogenous activatable gene, meaning that it should also be silent. This is indeed our observation. In the 129TF clones carrying a-transgenes, there is no detectable RT-PCR product for all but one of the genes tested (Figure 2A). For *Slc13A5*, only one of the three clones that we initially examined had expression, suggesting that expression in this clone might be due to the effect of integration site. Consistent with this, when three additional clones containing the *Slc13A5* BAC were examined, no *Slc13A5* expression was detected in any of them (Figure 2A). Given that RT-PCR primers were designed to equally amplify both the BAC transgenes and the endogenous genes, and given that these primers were validated on mouse and rat control cell types, the lack of RT-PCR product indicates that a-transgenes and their endogenous counterparts are both silent.

When several of the rat BACs carrying a-transgenes were introduced into C2C12 mouse myoblasts where the corresponding mouse genes were already expressed, both mouse and rat copies of the genes were expressed in every case, as determined by sequencing of RT-PCR products (data not shown). This indicates that the lack of expression of a-transgenes in 129TF is not due to technical artifacts, such as the BACs not carrying regulatory sequences required for expression, genetic incompatibility between species, or *de novo* silencing of these BACs after transfection. Taken together, these data confirm that activatable genes are indeed silent due to *trans*-acting mechanisms.

BAC transgenes corresponding to occluded endogenous genes are robustly expressed in most cases

Our expectations for o-transgenes were more complex. By definition, occluded genes are blocked by *cis*-acting chromatin mechanisms from responding to *trans*-acting factors, such that they remain silent irrespective of whether the cellular milieu is supportive or non-supportive of their expression. If the milieu is indeed supportive of the expression of an occluded gene, the corresponding BAC transgene is expected to be expressed, but if the milieu is not supportive, the BAC transgene should be silent. Of the 11 o-transgenes introduced into 129TF, 10 produced robust RT-PCR products, and for most of them, all clones showed positive expression (Figure 2B). This is in sharp contrast to the behavior of the a-transgenes (compare Figure 2A with 2B). Importantly, sequencing of the RT-PCR products revealed that expression came strictly from the BAC transgenes and not from their endogenous occluded counterparts (Figure 2B). It is important to note that the two a-transgenes that are silent in
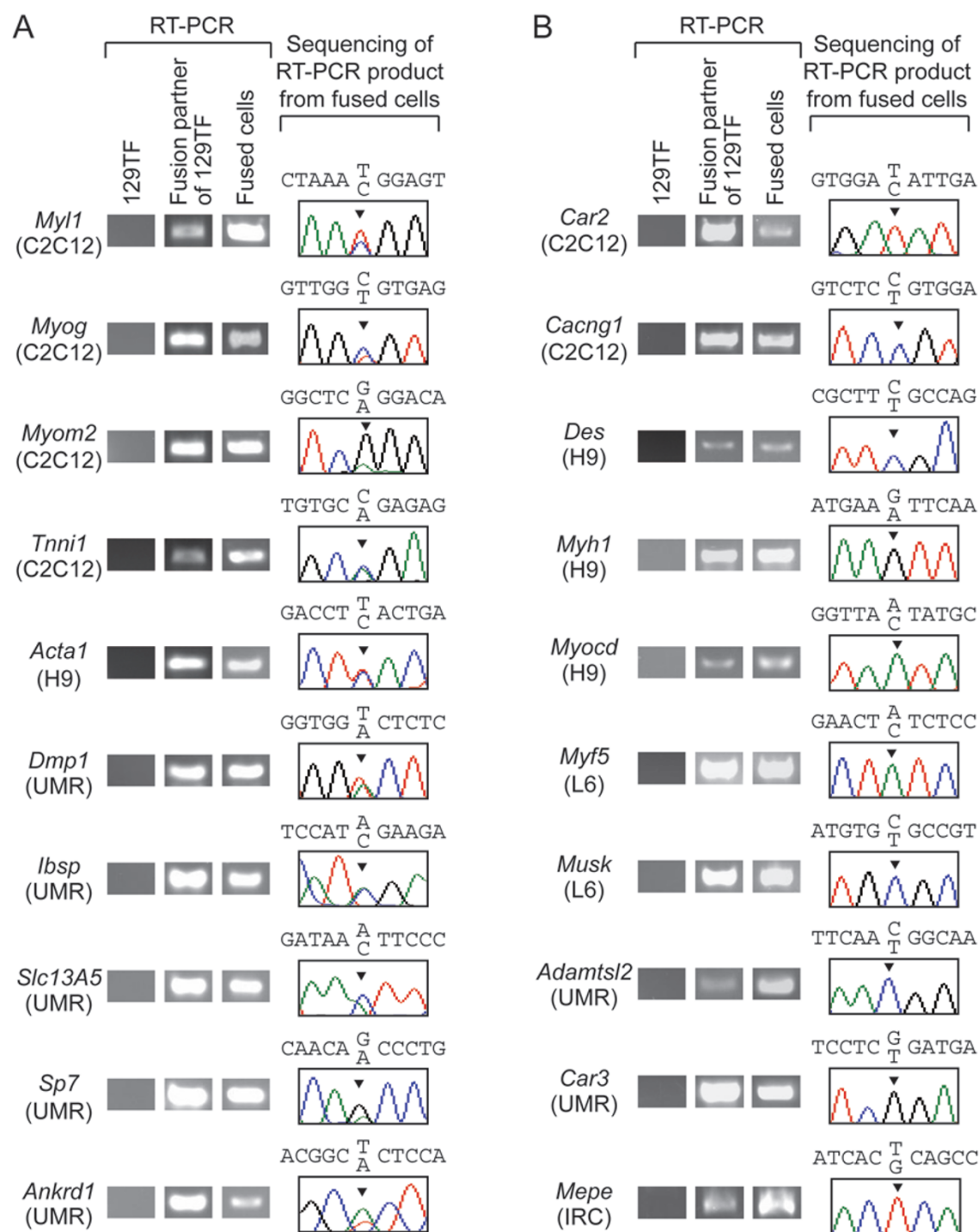
Figure 1 Assessing the activatable or occluded status of genes in 129TF. For each gene, the cell type used to fuse with 129TF in order to assess its activatable/occluded status in 129TF is indicated in parentheses below the gene name. Gel images are results of RT-PCR performed on 129TF, its fusion partner, and fused cells between the two. RT-PCR products from fused cells were sequenced, and the chromatograms corresponding to nucleotide sites that differ between 129TF and its fusion partner are shown. The sequence identities of these sites are given above the chromatograms (along with surrounding sequences), with the bottom and top alleles corresponding to 129TF and its fusion partner, respectively. Arrow heads indicate peaks of these sites in the chromatograms. **(A)** Activatable genes in 129TF. In fused cells, these genes show expression from both 129TF and its fusion partner. **(B)** Occluded genes in 129TF. In fused cells, these genes show no expression from 129TF and only expression from its fusion partner.
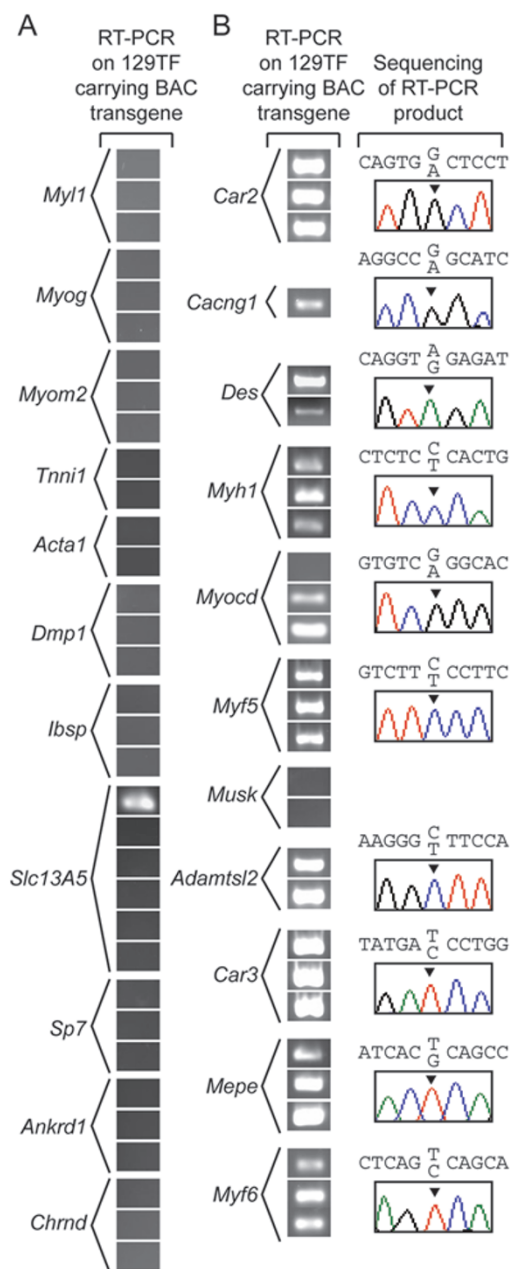
Figure 2 Expression status of BAC transgenes in 129TF. For each gene, RT-PCR was performed on all the 129TF clones carrying the corresponding BAC. **(A)** Lack of BAC transgene expression and endogenous gene expression in most 129TF clones carrying a-transgenes. **(B)** Robust BAC transgene expression in most 129TF clones carrying o-transgenes. For each o-transgene, RT-PCR products from all the clones were sequenced and a representative chromatogram corresponding to a nucleotide site that differs between the BAC transgene and the endogenous gene is shown. The sequence identities of such sites are given above the chromatograms (along with surrounding sequences), with the bottom and top alleles corresponding to the endogenous gene and the transgene, respectively. Arrow heads indicate peaks of these sites in the chromatograms. The chromatograms show that only the BAC transgenes and not the endogenous genes are expressed from the 129TF clones carrying o-transgenes.

129TF (*Dmp1* and *Ibsp*) and the one o-transgene that is active in 129TF (*Mepe*), all reside within the same BAC, arguing that the expression status of the transgenes is not dictated by some artifactual feature across the BAC. The sharply opposing behavior of the two categories of transgenes is statistically highly significant ($P < 5 \times 10^{-14}$ by Fisher's exact test performed at the level of clones or $P < 2 \times 10^{-5}$ when the test is performed at the level of genes). We were in fact quite surprised that the great majority o-transgenes produced robust expression. This argues that due to occlusion *in cis* solely, the endogenous genes are kept silent despite the presence of a cellular milieu that contains all the requisite factors *in trans* to turn them on.

Occlusion of Myf5 is essential in preventing fibroblasts from acquiring a muscle-like phenotype

To further examine the role of gene occlusion in cell fate restriction, we carried out detailed analyses of three of the genes, *Myf5*, *Myf6*, and *Myog*, used in the BAC analysis. All three are master regulators of skeletal myogenesis, as evidenced by the fact that their forced expression in certain non-muscle cell types leads to the manifestation of a muscle-like phenotype, such as the formation of myotubes [12-15]. Additionally, evidence suggests that *Myf5* and *Myf6* act upstream of *Myog* during myogenesis [16]. While all three genes are silent in 129TF, *Myf5* and *Myf6* are occluded, whereas *Myog* is activatable (Figure 1). *Myf5* and *Myf6* are adjacent in the genome and are present on a single BAC. We will refer to this rat BAC as the rMyf5/6 BAC and the BAC carrying rat *Myog* as the rMyog BAC. In all 129TF clones carrying the rMyf5/6 BAC, both *Myf5* and *Myf6* transgenes are robustly expressed, while the endogenous copies of *Myf5* and *Myf6* are silent (Figure 2B). In contrast, in all the 129TF clones carrying the rMyog BAC, the *Myog* transgene is silent just like the endogenous gene (Figure 2A).

Strikingly, myotubes were observed in the 129TF clones carrying the rMyf5/6 BAC (Figure 3A), indicating that the expression of the *Myf5* and/or *Myf6* transgenes have activated the myogenesis program in 129TF. RT-PCR confirmed that many muscle genes were indeed turned on in these cells (Figure 3B). Importantly, none of these downstream muscle genes are known to be occluded themselves, consistent with occluded genes remaining silent despite the presence of an activating cellular milieu. One of the genes activated in 129TF carrying the rMyf5/6 BAC is *Myog*, which is consistent with the known activatable status of this gene and the fact that *Myog* is thought to be functionally downstream of *Myf5* [16]. Thus, the introduction of the rMyf5/6 BAC alone into 129TF has triggered fibroblasts to acquire a muscle-
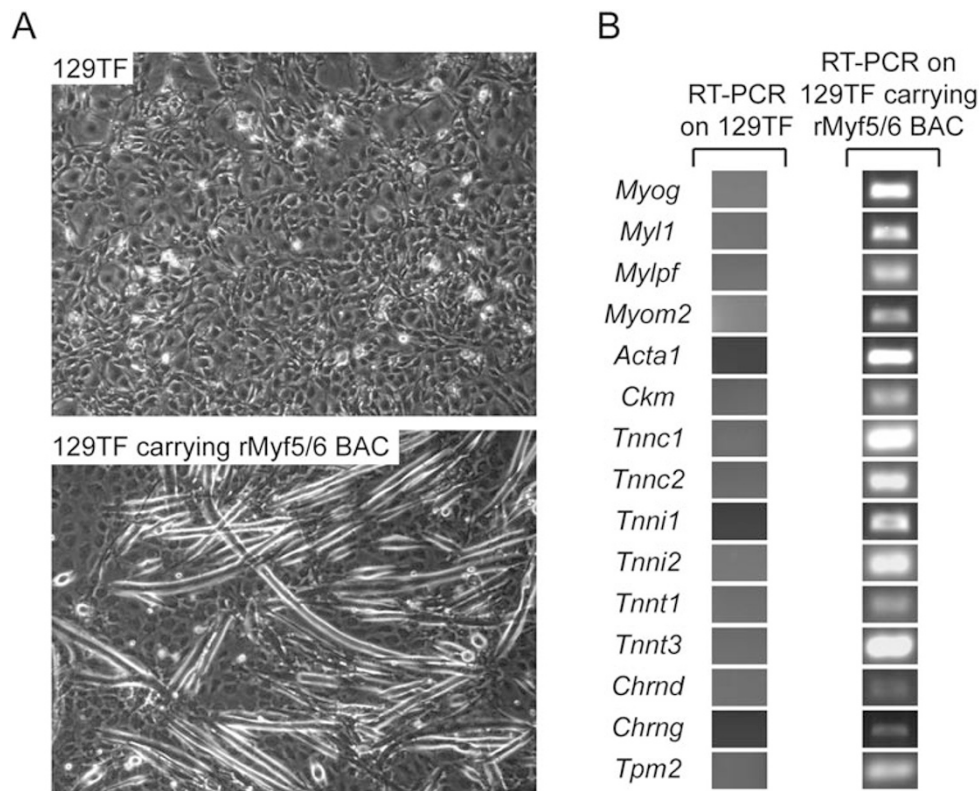
Figure 3 Acquisition of a muscle-like phenotype by 129TF carrying the rMyf5/6 BAC. **(A)** Phase-contrast microscopy images showing fibroblast morphology of regular 129TF and myotube morphology of 129TF carrying the rMyf5/6 BAC. **(B)** RT-PCR results showing that muscle-related genes are silent in normal 129TF and become activated in 129TF carrying the rMyf5/6 BAC.

like phenotype.

A simple interpretation of the robust expression of *Myf5* and *Myf6* transgenes in 129TF is that the cellular milieu of 129TF contains transcriptional activators for these two genes. However, there can be a more complex scenario, whereby, the 129TF milieu does not support the expression of one or both genes, but these genes engage in positive auto- and/or cross-regulation such that very low levels of initial background expression from the transgenes get amplified over time. To examine this possibility, we obtained a mouse BAC containing both *Myf5* and *Myf6* (hereon referred to as the mMyf5/6 BAC; see Supplementary information, Table S1), and knocked out both genes by replacing their first exons with a knockin cassette. When transfected into 129TF, robust expression of the knockin cassette was observed from both loci based on RT-PCR (data not shown). This argues that 129TF contains transcription factors capable of driving the expression of both *Myf5* and *Myf6*. We also knocked out *Myf5* or *Myf6* individually from the mMyf5/6 BAC and transfected the resulting BACs into 129TF. Myotube formation was observed only in the *Myf6* knockout BAC

(data not shown), indicating that the expression of the *Myf5* (but not *Myf6*) transgene alone can lead to myotube formation in 129TF.

BAC transgenes introduced into cells in early development behave similarly as endogenous genes

One inference from the above data is that the establishment of occlusion during differentiation and its subsequent maintenance in differentiated cells likely involve distinct mechanisms, the former being present only transiently during differentiation, whereas the latter remains in effect indefinitely. This would explain the observation that when a BAC transgene is introduced into fibroblasts in which the endogenous counterpart is already occluded, the transgene does not undergo *de novo* occlusion even after extended culture. Accordingly, we predicted that if the rMyf5/6 BAC is introduced into cells at an early stage of development before the occlusion of the endogenous *Myf5* and *Myf6* is established, the *Myf5* and *Myf6* transgenes should become occluded in fibroblasts just like the endogenous genes.

To test this, we performed pronuclear injection of

the rMyf5/6 BAC into one-cell mouse embryos to obtain transgenic animals carrying the BAC. We then derived tail fibroblasts from the animals, referred to as rMyf5/6(TG) fibroblasts. RT-PCR showed that both the *Myf5* transgene and the endogenous *Myf5* are silent in these fibroblasts, and the same is true for *Myf6* (Figure 4A). Furthermore, RT-PCR showed that the tissue expression patterns of *Myf5* and *Myf6* transgenes are the same as the endogenous genes, with strong expression in skeletal muscle for both genes and no detectable expression in a panel of other tissues (Figure 4A). Sequencing of the RT-PCR products from rMyf5/6(TG) muscle showed that both BAC transgene copies and endogenous copies of *Myf5* and *Myf6* are expressed (Figure 4A). To examine whether the BAC transgenes are occluded in rMyf5/6(TG) fibroblasts, we fused these cells with C2C12 mouse myoblasts and performed RT-PCR using mouse/rat common primers. Sequencing of RT-PCR products showed that for both *Myf5* and *Myf6*, only the mouse ortholog from the C2C12 genome is expressed in fused cells, indicating that the *Myf5* and *Myf6* BAC transgene is indeed occluded in rMyf5/6(TG) fibroblasts (Figure 4B).

For comparison, we also used pronuclear injection to produce transgenic animals carrying the rMyog BAC, and derived tail fibroblasts from transgenic animals. In these rMyog(TG) fibroblasts, RT-PCR revealed no expression from either the *Myog* transgene or the endogenous *Myog*. Furthermore, tissue expression patterns of the *Myog* transgene is the same as the endogenous gene, with prominent expression in skeletal muscle and no detectable expression in other tissues (Figure 4A). Sequencing of the RT-PCR product from muscle showed that both the BAC transgene copy and the endogenous copy of *Myog* are expressed (Figure 4A). To interrogate the occlusion status of the *Myog* transgene in rMyog(TG) fibroblasts, we fused these cells to C2C12, and performed RT-PCR on *Myog* using mouse/rat common primers followed by sequencing of RT-PCR product. This revealed that the *Myog* BAC transgene in rMyog(TG) fibroblasts is activatable, as it turned on in fused cells (Figure 4B).

Thus, the BAC transgene copies of *Myf5*, *Myf6* and *Myog* behave just like their endogenous counterparts in terms of expression patterns and occlusion status when introduced into cells at a very early developmental stage (i.e., one-cell embryo). However, when introduced acutely into terminally differentiated fibroblasts, the *Myog* transgene is silent just as its silent (and activatable) endogenous counterpart, whereas the *Myf5* and *Myf6* transgenes are actively expressed in contrast to their silent (and occluded) endogenous counterparts.

Taken together, these observations argue that the occlusion of *Myf5* and *Myf6* is established early during development and subsequently maintained in differ-
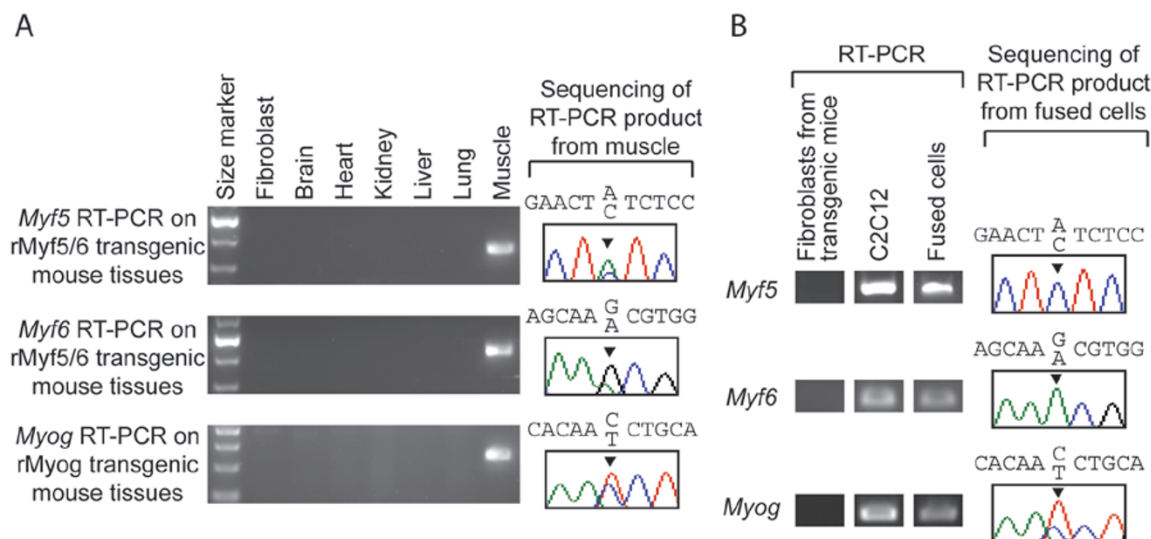

Figure 4 Expression pattern and occlusion status of *Myf5*, *Myf6*, and *Myog* transgenes in transgenic mice. **(A)** Comparable expression pattern of transgene copies and endogenous copies of *Myf5*, *Myf6*, and *Myog* in multiple tissues and tail fibroblasts. RT-PCR using mouse/rat common primers revealed robust expression only in the skeletal muscle for *Myf5*, *Myf6*, and *Myog*. For each gene, sequencing of the RT-PCR product from muscle revealed expression from both the transgene and the endogenous gene. **(B)** Different occlusion status of the *Myf5* and *Myf6* transgenes versus the *Myog* transgene. In fused cells between fibroblasts derived from transgenic mice and C2C12, the *Myf5* and *Myf6* transgenes are silent, indicating that they are occluded. In contrast, the *Myog* transgene is expressed, indicating that it is activatable.

entiated cells, and that the occlusion of these genes by *cis*-acting mechanisms is absolutely required for their stable silencing in a cellular milieu that would otherwise activate them and cause cells to acquire a muscle-like phenotype. In contrast, *Myog* does not undergo occlusion in fibroblasts, and its silent state in these cells is due to *trans*-acting mechanisms (i.e., absence of transcriptional activators). The fact that BAC transgenes in transgenic animals behave in the same way as their endogenous counterparts also indicates that these BACs contain all the necessary regulatory sequences for their proper expression (and proper occlusion in the case of the rMyf5/6 BAC).

Discussion

To understand cell fate restriction is to understand how gene expression patterns characteristic of individual cell types are stably maintained. While both *trans*-acting mechanisms involving diffusible factors and *cis*-acting mechanisms involving chromatin modifications have been implicated [3-7], definitive data on the nature and extent by which either *trans* or *cis* mechanisms contribute to cell type-specific gene expression patterns are lacking. The present study provides compelling evidence that gene occlusion – a mode of gene silencing mediated by *cis* mechanisms – plays a prominent and critical role in cell fate restriction.

By using a BAC transgene approach, we make the surprising observation that for the majority of occluded genes, the transcriptional milieu *in trans* is actually supportive of their expression, meaning that the cell contains transcriptional activators for most occluded genes. Our study further shows that without the occlusion of a single gene, *Myf5*, fibroblasts would acquire a muscle-like phenotype. Our recent effort to systematically map occluded genes showed that the number of occluded genes in a typical somatic cell type could be in the thousands (manuscript submitted). The present study argues that most of these genes would have been expressed if it were not for their occluded status. We therefore conclude that occlusion has an essential function in safeguarding somatic cell fate by preventing the expression of a large number of lineage-inappropriate genes.

The lack of expression of the single o-transgene, *Musk*, may be due to the absence of a transcriptionally supportive milieu, suggesting that the presence of such a milieu is not completely universal for all occluded genes, even if it is the case for the majority. However, an alternative explanation is that the lack of expression of this outlier gene is due to the fact that it is the largest gene included in this study, and the relatively small amount of flanking genomic sequence included in the BAC may have resulted in the loss of a necessary positive regulatory element far from the gene.

The importance of occlusion in cell fate restriction, as reported here, would seem incompatible with the relative ease by which somatic cells can be dedifferentiated towards pluripotency by experimental manipulations, such as SCNT into oocytes [17, 18], or forced expression of the iPSC factors [19, 20]. In a submitted manuscript, we resolved this paradox by showing that pluripotent stem cells, unlike somatic cells, have the unique capacity for global deocclusion (i.e., erasure of occlusion on a whole-genome scale). We therefore argue that somatic cells can be experimentally reprogrammed to pluripotency either by tapping into the existing deocclusion capacity of pluripotent cells (i.e., SCNT into oocytes) or by recapitulating key aspects of the deocclusion machinery (i.e., the use of defined factors to create iPSCs).

The fact that occluded genes in somatic cells do not turn on, even though transcriptional activators for these genes are often present in the cellular milieu, would seem inconsistent with reports that forced expression of certain transcription factors can drive one somatic cell type to transdifferentiate into another, such as fibroblasts to myotube-like cells [21], to hepatocytes-like cells [22, 23], to cardiomyocyte-like cells [24, 25], to macrophage-like cells [26], and to neuron-like cells [27]. Indeed, one influential school of thought holds that the differentiated state is continuously and dynamically regulated by the balance of transcription factors, and a tip in such balance could lead to transdifferentiation [3, 4, 28, 29]. We offer two possible explanations for the apparent contradiction between occlusion and transdifferentiation. First, previous claims of transdifferentiation by forced expression of transcription factors often examined only a limited number of cell type-specific markers. It is hence possible that the downstream genes activated by these transcription factors are mostly activatable genes and many – if not all – occluded genes are still silent due to occlusion. If this is the case, then some of the claims of transdifferentiation might not entail the complete conversion of one cell type to another; but rather, they might represent a partial (and incomplete) acquisition of gene expression patterns characteristic of the second cell type, a situation that is perhaps better called "ectopic differentiation" rather than true transdifferentiation [10]. This is supported by our observation that when the rMyf5/6 BAC is transfected into fibroblasts, although the *Myf5* BAC transgene is expressed and this expression induces the cells to assume a muscle-like phenotype, these cells do not express the endogenous *Myf5* (which is occluded) or other occluded genes examined. Likewise, forced overexpression

of the related bHLH transcription factor *MyoD1* as a transgene in 129TF was also able to induce muscle-like phenotype without activation of endogenous *Myf5* (unpublished data). Thus, whereas these cells do show some resemblance to muscle in terms of gene expression and cellular physiology, they are different from naturally derived muscle cells (e.g., they do not express endogenous *Myf5*). Indeed, for the "transdifferentiation" studies cited above, the endogenous genes in 129TF corresponding to many of the overexpressed transgenes were found to be occluded (manuscript submitted), prompting us to speculate that the endogenous copies of these genes (as well as other occluded genes) might actually not be activated in these transdifferentiation studies. Notwithstanding the above discussion, we also propose an alternative explanation, namely, it is also possible that the overexpression of certain transcription factors far beyond their physiological levels might indeed be able to activate some occluded downstream genes and might even permanently erase occlusion. This is plausible, given the fact that many transdifferentiation experiments use strong promoters to drive exogenously introduced transcription factor genes, which can result in supraphysiological expression levels.

Importantly, our results suggest that the transcriptional environment *in trans* in different cell types are perhaps much more similar than previously appreciated, and that the differential expression of genes between different cell types is, to a large extent, the result of differential occlusion *in cis* rather than the presence of different factors *in trans*. Indeed, we propose that at the most fundamental level, the identities of individual cell types are defined largely by the occlusion status of their genomes.

Materials and Methods

Ethics statement

All animal work was conducted according to relevant national and international guidelines.

Cell culture

The C2C12 mouse myoblasts of C3H strain background were obtained from ATCC (Cat# CRL-1772), and two clonal populations, named C2C12-GPc60 and C2C12-RHc2, were derived from cells transduced with a lentiviral vector containing EGFP driven by the human EF1A promoter and puromycin resistance, and cells transduced with a lentiviral vector containing dTomato [30] driven by the human EF1A promoter and hygromycin resistance. Tail fibroblasts derived from rMyf5/6(TG) and rMyog(TG) mice were transduced with a lentiviral vector containing dTomato driven by the human EF1A promoter and hygromycin resistance. Lentiviral vectors are as described previously [31]. The origins of the rat cells are as follows: IRC (full name: IRC-RHc17) from IRC chondrocytes [32] (gift from Walter Horton), L6 (full name: L6-RHc6) from L6 myoblasts (ATCC, Cat# CRL-1458), H9 (full

name: H9-RHcA10) from H9c2(2-1) cardiomyoblasts (ATCC, Cat# CRL-1446), and UMR (full name: UMR-RHc7) from UMR-106 osteosarcoma (ATCC, Cat# CRL1661). Each of these rat cell types was a clonal population derived from cells transduced with a lentiviral vector carrying constitutively expressed DsRed-Express2 [33] or dTomato driven by the human EF1A promoter and hygromycin resistance as described previously. They were cultured under conditions as published or recommended by the vendor. The lentivirus-transduced cells used in this study are available through Cyagen Biosciences. Fusions between tail fibroblasts from BAC transgenic mice and C2C12-GPc60, as well as between 129TF and C2C12-RHc2, were carried out as follows. C2C12 cells were first cultured under low-serum conditions. The two cell types were then plated together for 6 h, followed by overlaying a solution of 37 °C PEG 1500 MW (50% w/w in serum-free DMEM) for 1 min. After removal of PEG, cells were washed three times with warm serum-free DMEM and allowed to recover for 4 h at 37 °C, followed by transition to media containing 1 μg/ml puromycin and 400 μg/ml hygromycin to select for dual drug-resistant fused cells. Total RNA was harvested for cDNA preparation 5 days after fusion. All other fusions were performed following the same general protocol as follows. Cells were plated together for at least 2 h (or overnight). Prior to fusion, cells were washed with serum-free DMEM, and PEG 1500 MW or 1000 MW (50% w/v in serum-free DMEM) was added for 1 min. After removal of PEG, cells were washed three times with serum-free DMEM and allowed to recover for 2 h. Following this, cells were split to a lower density and plated in media containing both puromycin and hygromycin to select for double drug-resistant fused cells. Daily trypsinization aided in selection and purification of fused cells, with total RNA harvested between 6-7 days post fusion. Variations in this protocol included cell to cell ratio, cell density, and concentrations of puromycin and hygromycin, all of which were determined empirically for each fusion. L6 was differentiated under low-serum condition prior to fusion.

RT-PCR-Seq

Total RNA was purified from cell or tissue samples using the ZR RNA MiniPrep kit (Zymo Research) and treated with DNase I to remove contaminating genomic DNA. Reverse transcription was carried out using the SuperScript III First-Strand Synthesis System (Invitrogen) with random hexamer primers following vendor's instructions. PCR was performed in 20 μl reactions and 35 cycles using Platinum Taq Hi-Fidelity (Invitrogen) following vendor's instructions. Each reaction used template cDNA converted from 2.5 ng of total RNA. PCR products were treated with shrimp alkaline phosphatase and *E. coli* exonuclease I (US Biologicals), followed by an ABI BigDye Terminator reaction and sequencing using an ABI 3730xl DNA Analyzer (Applied Biosystems). Mouse/rat (or mouse/mouse) common primers were designed from primer-binding sequences that are identical between the two species (or between the two mouse strains being investigated), and which flank sequence differences between the two, without indels, based on cDNA sequence alignments. Whenever possible, primers were designed such that the resulting amplicon spanned one or more introns. This way, contaminating genomic DNA would not result in RT-PCR product of the right size. Mouse and rat cDNA sequences were based on RefSeq sequences retrieved from NCBI. Polymorphisms between mouse strains were identified from the

Perlegen Mouse HapMap as published [34], or by genomic PCR and sequencing of PCR product. To perform genomic PCR and sequencing, primers were designed to amplify individual exons from genomic DNA plus at least 50 bp flanking sequences on either side, followed by sequencing of PCR products. Sequences of PCR primers are available upon request. All RT-PCR primers were validated based on their ability to amplify products from positive control cell types where the target genes were known to be expressed.

BAC preparation and transfection

BACs of interest were purchased from the Children's Hospital Oakland Research Institute BACPAC Resources Center, and were modified to introduce a neomycin resistance cassette as described previously [35]. Gene replacement in the mMyf5/6 BAC was carried out using Red/ET recombination system (Gene Bridges) following vendor's instructions and as previously described [36]. For either *Myf5* or *Myf6*, the ORF portion of the first exon plus two adjacent bases of the first intron were replaced with a knockin cassette [36]. *Myf5* was replaced with a cassette consisting of dTomato, followed by IRES (internal ribosome entry site), GB3 (a bacterial promoter), and *neo* (neomycin resistance gene). *Myf6* was replaced with a cassette consisting of eGFP, followed by IRES, GB3, and *bsd* (blasticidin resistance gene). BACs were prepared using the PowerPrep HP Maxiprep kit (Marligen Biosciences), and transfected as described [37]. 129TF cells were plated at approximately 30% confluency in 30-mm dishes the day before transfection, expanded to 10-cm dishes 2 days after transfection, and maintained in 400 μg/ml neomycin (G418) selection medium 5 days after transfection. For isolation of stable BAC transfected clones, neomycin-resistant cells were trypsinized, and individual cells distributed into wells of 96-well plates using a FACSAria II Flow Cytometer (BD Biosciences). Clones were expanded in medium containing 400 μg/ml neomycin until subsequent analysis.

Generation of BAC transgenic mice

The production of transgenic mice and the derivation of tail fibroblasts from transgene-positive animals were performed by Cyagen Biosciences. Unmodified rMyf5/6 BAC and rMyog BAC were used to make transgenic mice by pronuclear injection of BAC DNA into one-cell embryos of FVB strain background following standard procedures.

Acknowledgments

References

1   Waddington CH. Organisers and Genes. London: Cambridge University Press, 1940.

2   Waddington CH. The Strategy of the Genes. London: George Allen & Unwin, 1957.

3   Blau HM, Baltimore D. Differentiation requires continuous regulation. *J Cell Biol* 1991; **112**:781-783.

4   Graf T, Enver T. Forcing cells to change lineages. *Nature* 2009; **462**:587-594.

5   Yamanaka S, Blau HM. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* 2010; **465**:704-712.

6   Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 2007; **447**:425-432.

7   Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. *Cell* 2007; **128**:635-638.

8   Hirabayashi Y, Gotoh Y. Epigenetic control of neural precursor cell fate during development. *Nat Rev Neuroscience* 2010; **11**:377-388.

9   Lee JH, Bugarija B, Millan EJ, *et al.* Systematic identification of *cis*-silenced genes by *trans* complementation. *Hum Mol Genet* 2009; **18**:835-846.

10   Lahn BT. The "occlusis" model of cell fate restriction. *BioEssays* 2010; **33**:13-20.

11   Lee JH, Gaetz J, Bugarija B, *et al.* Chromatin analysis of occluded genes. *Hum Mol Genet* 2009; **18**:2567-2574.

12   Braun T, Buschhausen-Denker G, Bober E, Tannich E, Arnold HH. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J* 1989; **8**:701-709.

13   Wright WE, Sassoon DA, Lin VK. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell* 1989; **56**:607-617.

14   Braun T, Bober E, Winter B, Rosenthal N, Arnold HH. Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *EMBO J* 1990; **9**:821-831.

15   Miner JH, Wold B. Herculin, a fourth member of the MyoD family of myogenic regulatory genes. *Proc Natl Acad Sci USA* 1990; **87**:1089-1093.

16   Tapscott SJ. The circuitry of a master switch: MYOD and the regulation of skeletal muscle gene transcription. *Development* 2005; **132**:2685-2695.

17   Gurdon JB. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J Embryol Exp Morphol* 1962; **10**:622-640.

18   Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KH. Viable offspring derived from fetal and adult mammalian cells. *Nature* 1997; **385**:810-813.

19   Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; **126**:663-676.

20   Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007; **448**:313-317.

21   Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 1987; **51**:987-1000.

22   Huang P, He Z, Ji S, *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* 2011; **475**:386-389.

23   Sekiya S, Suzuki A. Direct conversion of mouse fibroblasts

to hepatocyte-like cells by defined factors. *Nature* 2011; **475**:390-393.

24 Ieda M, Fu JD, Delgado-Olguin P, *et al.* Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* 2010; **142**:375-386.

25 Efe JA, Hilcove S, Kim J, *et al.* Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat Cell Biol* 2011; **13**:215-222.

26 Feng R, Desbordes SC, Xie H, *et al.* PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proc Natl Acad Sci USA* 2008; **105**:6057-6062.

27 Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 2010; **463**:1035-1041.

28 Blau HM. How fixed is the differentiated state? Lessons from heterokaryons. *Trends Genet* 1989; **5**:268-272.

29 Slack JM. Metaplasia and transdifferentiation: from pure biology to the clinic. *Nat Rev Mol Cell Biol* 2007; **8**:369-378.

30 Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp.* red fluorescent protein. *Nat Biotechnol* 2004; **22**:1567-1572.

31 Qin JY, Zhang L, Clift KL, *et al.* Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* 2010; **5**:e10611.

32 Horton WE Jr, Cleveland J, Rapp U, *et al.* An established rat cell line expressing chondrocyte properties. *Exp Cell Res* 1988; **178**:457-468.

33 Strack RL, Strongin DE, Bhattacharyya D, *et al.* A noncytotoxic DsRed variant for whole-cell labeling. *Nat Methods* 2008; **5**:955-957.

34 Frazer KA, Eskin E, Kang HM, *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 2007; **448**:1050-1053.

35 Wang Z, Engler P, Longacre A, Storb U. An efficient method for high-fidelity BAC/PAC retrofitting with a selectable marker for mammalian cell transfection. *Genome Res* 2001; **11**:137-142.

36 Poser I, Sarov M, Hutchins JR, *et al.* BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat Methods* 2008; **5**:409-415.

37 Phelps SF, Illenye S, Heintz NH. Manipulating genes and gene copy number by bacterial artificial chromosomes transfection. *Methods Mol Biol* 2007; **383**:153-163.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)