Human Molecular Genetics, 2009, Vol. 18, No. 5 **835–846**
doi:10.1093/hmg/ddn409
Advance Access published on December 2, 2008

Systematic identification of *cis*-silenced genes by *trans* complementation

Jae Hyun Lee[1], Branimir Bugarija[1], Enrique J. Millan[1], Noah M. Walton[1], Jedidiah Gaetz[1], Croydon J. Fernandes[1], Wei-Hua Yu[2], Nitzan Mekel-Bobrov[1], Tammy W. Vallender[1], Gregory E. Snyder[1], Andy Peng Xiang[2] and Bruce T. Lahn[1,2,*]

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

Received November 6, 2008; Revised and Accepted December 1, 2008

A gene's transcriptional output is the combined product of two inputs: diffusible factors in the cellular milieu acting in *trans*, and chromatin state acting in *cis*. Here, we describe a strategy for dissecting the relative contribution of *cis* versus *trans* mechanisms to gene regulation. Referred to as *trans* complementation, it entails fusing two disparate cell types and searching for genes differentially expressed between the two genomes of fused cells. Any differential expression can be causally attributed to *cis* mechanisms because the two genomes of fused cells share a single homogenized milieu in *trans*. This assay uncovered a state of transcriptional competency that we termed 'occluded' whereby affected genes are silenced by *cis*-acting mechanisms in a manner that blocks them from responding to the *trans*-acting milieu of the cell. Importantly, occluded genes in a given cell type tend to include master triggers of alternative cell fates. Furthermore, the occluded state is maintained during cell division and is extraordinarily stable under a wide range of physiological conditions. These results support the model that the occlusion of lineage-inappropriate genes is a key mechanism of cell fate restriction. The identification of occluded genes by our assay provides a hitherto unavailable functional readout of chromatin state that is distinct from and complementary to gene expression status.

INTRODUCTION

Multicellular life is defined by the presence, within a single organism, of a wide array of cell types bearing the same genome but disparate physiological functions. This is typically achieved through the progressive differentiation of multipotent stem cells into functionally specialized cells. As a general rule, differentiated cell types can stably maintain their phenotypic identities despite fluctuations in extracellular environment and intracellular regulatory networks (1). How cell type identity is maintained at the molecular level is a central but poorly understood question in biology. One attractive idea is that the phenotypic identity of differentiated cells is maintained via the silencing of lineage-inappropriate genes—i.e. genes promoting alternative lineages which, if expressed aberrantly, would lead to the manifestation of incorrect cellular phenotypes (2–5).

This idea is in line with the increasing recognition that the transcriptional output of a gene is the combined product of two distinct inputs. The first is the *trans*-acting milieu of the cell, defined as all the diffusible factors that collectively impinge on gene regulatory sequences to promote or repress expression. The second is the *cis*-acting chromatin state of the gene itself, defined as the full complement of chromatin marks at the locus such as DNA methylation, histone modifications and the binding of chromatin remodeling factors, which in combination determine how the locus responds to its milieu. Numerous studies have found that particular chromatin marks such as DNA methylation and histone hypoacetylation are enriched at silent loci of the genome (6–13). In most cases, however, the exact contribution of these chromatin marks to the silent state cannot be teased apart from the contribution of milieu. This is because it is difficult to know

*To whom correspondence should be addressed. Tel: +1 7738344393; Fax: +1 7737020271; Email: blahn@bsd.uchicago.edu

Table 1. Expression patterns of occluded, transactivated and extinguished genes

Expression pattern in reprogrammer		Expression pattern in responder		Conclusion
Before fusion	After fusion	Before fusion	After fusion	
Active	Active	Silent	Silent	Gene in responder occluded
Active	Active	Silent	Active	Gene in responder transactivated and hence competent
Active	Silent	Silent	Silent	Gene in reprogrammer extinguished

whether chromatin marks at silent loci are the cause or consequence of silencing, or to what extent is the silent status of a gene and its associated chromatin marks reversible when cellular milieu changes (9,14). As such, whether gene silencing by chromatin-based *cis* mechanisms plays a key role in restricting cell fate remains to be resolved.

Monoallelic silencing such as X inactivation and imprinting is a clear exception to the above ambiguity. Here, it can be unequivocally ascertained that silencing is due to *cis*-acting chromatin mechanisms in a manner independent of milieu. The hallmark of monoallelic silencing is the differential expression of two copies of a gene—one silent and one active—in the same cell (15–17). The active copy serves as a positive control, attesting to the presence of a milieu that is conducive to the expression of the gene. In this context, the silent copy, which is bathed in the same milieu, must have been blocked from the milieu's action by the *cis* effect of its chromatin state. Thus, at least in the case of monoallelic silencing, the transcriptional competency of a gene can be defined as existing in either of two states. One is the 'competent' state whereby a gene is capable of responding to the milieu of the cell, such that it is active if appropriate transcription activators are present, and silent if activators are absent or repressors are present. The other can be called the 'occluded' state whereby a gene is no longer capable of responding to the cell's milieu and remains silent even in the presence of a transcriptionally conducive milieu.

It is reasonable to hypothesize that during development, some genes might become biallelically occluded by mechanisms similar to monoallelic silencing, and that this process could play an essential role in maintaining the phenotypic identities of cells. A key test of this model is the identification of biallelically occluded genes. However, the lack of a positive control—the equivalent of the active copies for monoallelically silenced genes—poses a technical challenge in ascertaining the presence of biallelically occluded genes. This is because it is impossible, without such a control, to definitively differentiate whether a silent gene is in the occluded state or whether it is competent but not expressed simply owing to the lack of a conducive milieu. Furthermore, biochemical modifications of chromatin, which regulate gene expression in *cis*, are immensely complex (for example, there are over 100 known chromatin marks) (18), thus limiting the use of a 'bottom up' approach to differentiate *cis* versus *trans* regulation.

Here, we describe the *trans* complementation assay, which allows the systematic identification of biallelically occluded genes. The approach is to fuse two disparate cell types and search in fused cells for genes silent in one genome but active in the other. Similar to monoallelic silencing, the active copies of genes serve as a positive control, with which the occluded state of the silent copies can be ascertained.

RESULTS

Identification of occluded genes via cell fusion

To identify occluded genes within specific cell types, we employed a cell fusion strategy. For ease of description, one of the two cell types being fused will be referred to as the responder and the other the reprogrammer. The goal is to identify occluded genes in the responder, which are defined operationally as genes silent in the responder genome of fused cells but active in the reprogrammer genome of the same cells (Table 1). We chose human lung fibroblasts (hereon abbreviated hLF) as the responder and mouse skeletal muscle myoblasts (mSMM) as the reprogrammer. Using cells from different species, sequence divergence between orthologs can be exploited to distinguish whether a transcript in fused cells is produced from the reprogrammer genome or the responder genome.

The two cell populations were labeled by different fluorescent dyes and fused by polyethylene glycol. Dual fluorescent cells, which represent a small fraction of the total, were isolated by fluorescence-activated cell sorting (FACS) (Supplementary Material, Fig. S1A). Microscopy confirmed that FACS-isolated cells were predominantly (>98%) fusions between hLF and mSMM, as they contained multiple nuclei of two distinct morphologies (hLF nuclei are larger and have weaker DAPI staining relative to mSMM) (Supplementary Material, Fig. S1B). For a subset of experiments, cells of heterotypic fusion (i.e. fusion between hLF and mSMM) were further enriched by antibiotics that eliminated unfused cells or cells of homotypic fusion. Of the fused cells, >70% showed equal numbers of hLF versus mSMM nuclei, the great majority of which possessed one hLF and one mSMM nucleus, whereas the rest contained two hLF and two mSMM nuclei. Less than 30% of cells showed unequal numbers of hLF and mSMM nuclei, the majority of which had an overrepresentation of mSMM nuclei (Supplementary Material, Fig. S1C). Fused cells were cultured for varying periods of time to allow for the resetting of gene expression in the new cellular milieu. Regardless of culture period and medium formulation, fused cells remained as multinucleated heterokaryons, indicating that they had lost the ability to divide after fusion. We found that gene expression patterns became stabilized within 3 days of fusion (see what follows). We therefore focused on day 4 post-fusion for our analysis of gene expression.

To interrogate gene expression in hLF and mSMM before and after fusion, we used human and mouse Affymetrix microarrays. Although there is significant sequence divergence between human and mouse genomes (average 16% in coding regions), a human transcript in fused cells may still hybridize to orthologous probes on the mouse arrays and vice versa, given that the arrays are not designed for species-specific hybridization. To examine how serious a problem cross-species hybridization might be, we hybridized cRNA from

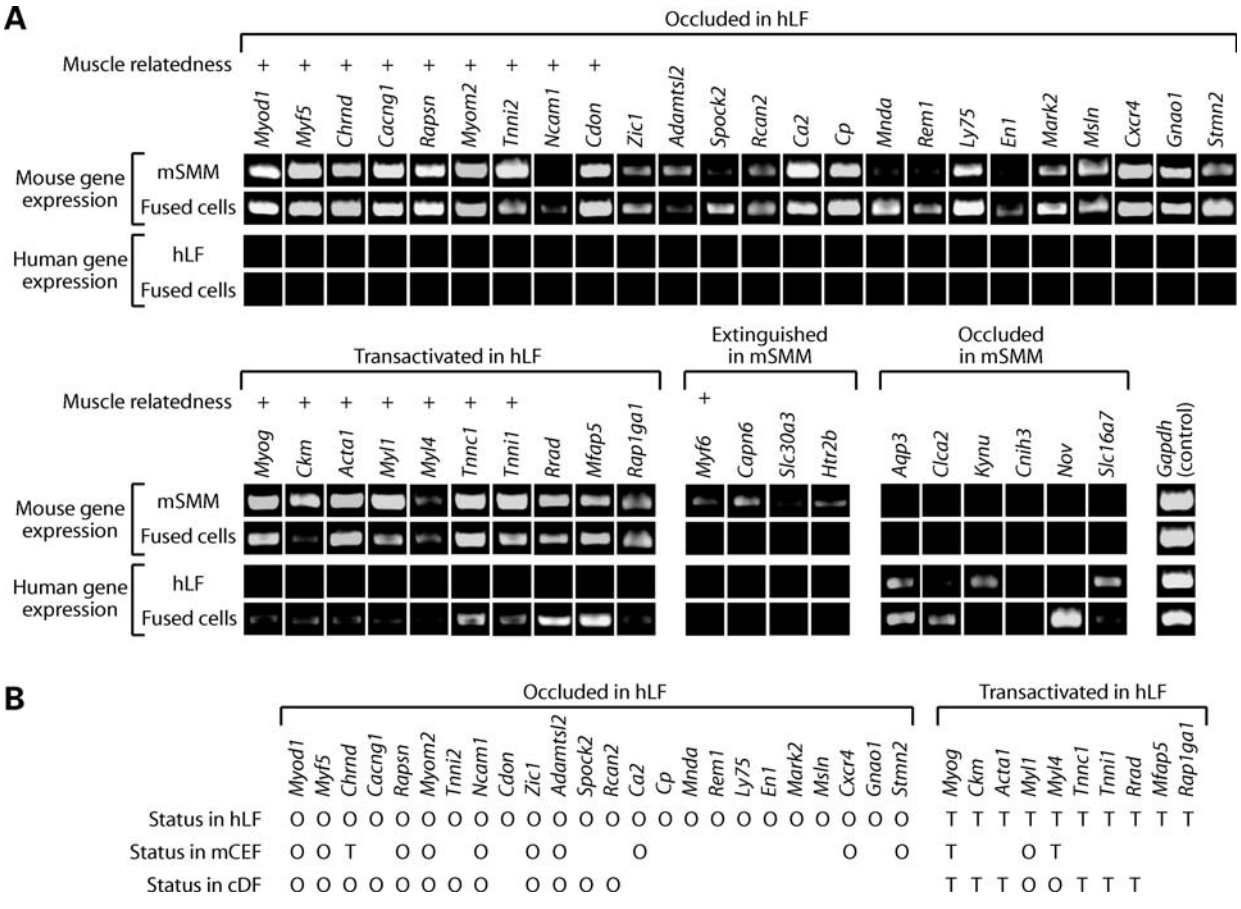

Figure 1. Identification of occluded genes. (**A**) Expression analysis of the hLF–mSMM fusion by RT–PCR. Four classes of genes are shown: occluded hLF genes, transactivated hLF genes, extinguished mSMM genes and occluded mSMM genes. For each gene, four RT–PCR results are shown: the two on the top target the mSMM ortholog before and after fusion, whereas the two on the bottom target the hLF ortholog before and after fusion. Known muscle-related genes are indicated by '+' above the gene name. (**B**) Summary of whether genes are occluded (denoted by O) or transactivated (T) in hLF, mCEF and cDF. hLF, human lung fibroblasts; mSMM, mouse skeletal muscle myoblasts; mCEF, mouse cloned embryonic fibroblasts; cDF, chimpanzee dermal fibroblasts.

each cell type to both the human and the mouse arrays. When cRNA from the correct species was hybridized to the arrays, ~45% of all the genes were called present. In contrast, cRNA from the wrong species only led to ~10% of the genes being called present. This shows that the arrays have sufficient species specificity to interrogate the expression of a considerable fraction of genes in fused cells, but it also demonstrates the need for validation by more stringent methods.

Four sets of array data were generated: hLF on human arrays, mSMM on mouse arrays, fused cells on human arrays and fused cells on mouse arrays. To ensure robustness of the analysis, we first narrowed down to a list of genes shown by the array data to be active in mSMM but silent in hLF prior to fusion. If, in the fused cells, these genes remain active in the mSMM genome and silent in the hLF genome, they would be placed in a candidate list of occluded hLF genes.

This analysis generated a candidate list of 279 putatively occluded hLF genes, all of which were subject to RT–PCR validation. For each gene, we designed and confirmed mouse-specific and human-specific RT–PCR primers to allow mouse and human gene expression in fused cells to be interrogated independently. Consistent with previous Affymetrix microarray studies (19), our RT–PCR analysis showed that absence calls in the array data are much less reliable than presence calls. As a result, a large number of the candidate-occluded hLF genes from the array data were shown by RT–PCR to be expressed at appreciable levels in hLF both before and after fusion. Winnowing out these and other false leads, 24 genes were confirmed by RT–PCR to exhibit expression patterns consistent with their occluded status in hLF (Fig. 1A and Supplementary Material, Table S1). Of these, nine have known muscle-related functions (indicated in Fig. 1A). PCR on genomic DNA of fused cells using human-specific primers successfully amplified the hLF copies of all these genes, indicating that their lack of expression in fused cells is not due to the absence of hLF chromosomes (data not shown). Indeed, it is unlikely that chromosome loss should occur in mitotically arrested heterokaryons.

Applying the criteria in Table 1, we also obtained a candidate list of 1040 putatively transactivated hLF genes. Given that transactivation was not the focus of the study, we selected only a subset of 202 genes for RT–PCR validation.

Many genes failed validation because RT–PCR detected appreciable levels of expression in hLF both before and after fusion. For a lot of these, RT–PCR did show increased expression after fusion, but we did not consider them as transactivated genes per our stringent definition of transactivation. This led to the identification of 10 transactivated hLF genes, of which 7 have known muscle-related functions (Fig. 1A and Supplementary Material, Table S1). For three of the genes (*Ckm*, *Acta1* and *Myl1*), their transactivation is consistent with previous reports (20,21). For transactivated genes that showed significantly less amplification of the hLF transcripts than the mSMM transcripts in fused cells (such as *Myog*, *Acta1* and *Rap1ga1*), additional sets of primers confirmed that the differences in amplification reflected actual gene expression differences between the human and mouse orthologs, rather than differences in PCR efficiencies. In principle, the observed differences in gene expression between human and mouse transcripts in fused cells could be due to at least three possibilities. First, these genes could be partially occluded such that they turn on in response to the introduction of a conducive milieu, but not to the full extent possible (see Discussion). Second, the hLF cell population may be heterogeneous, with the genes in question being occluded in some but not all the cells. A third possibility is that incompatibility between mouse transcription factors made from the mSMM genome and human *cis*-regulatory sequences in the hLF genome results in only partial activation of these genes. This possibility is addressed in greater detail later.

Our results demonstrate that, although both occluded and transactivated genes are silent in hLF prior to fusion, they clearly exist in two distinct states of transcriptional competency. Occluded genes do not become active even in the presence of a transcriptionally conducive milieu. In contrast, transactivated genes exist in a competent (though inactive) state that can turn on in response to the introduction of *trans*-acting factors in the milieu.

Ad hoc RT–PCR analysis also uncovered four extinguished mSMM genes and six occluded mSMM genes (Fig. 1A). Extinction could result either from the introduction of transcriptional repressors or from the dilution or disappearance of transcriptional activators upon fusion (22). For extinguished mSMM genes, it is not possible to determine whether their orthologs in hLF are occluded or not. The presence of occluded mSMM genes indicates that a given cell fusion experiment can be used to identify occluded genes in both fusion partners, even though we chose to focus only on hLF in this study.

The identification of occluded hLF genes was carried out in a systematic and unbiased fashion, in the sense that all the candidate occluded genes based on the array data were subject to RT–PCR validation. The final tally of 24 occluded hLF genes therefore likely represents a considerable fraction of all occluded hLF genes in the context of the hLF–mSMM fusion experiment. In contrast, the transactivated hLF genes, extinguished mSMM genes and occluded mSMM genes were uncovered by less systematic means.

The specification of the myogenic lineage is controlled by four transcription factors, *Myod1*, *Myf5*, *Myog* and *Myf6* (23–27). Of these myogenic master triggers, *Myod1* and *Myf5* are occluded in hLF, *Myog* is transactivated (and therefore competent) in hLF and *Myf6* is extinguished in mSMM (and therefore may be either occluded or competent in hLF). Interestingly, *Myod1* and *Myf5* are known to be upstream of *Myog* and *Myf6* in driving myogenic programs, and they also engage in positive autoregulation and positive cross-regulation (28,29). Given such a regulatory circuit, should *Myod1* and *Myf5* not undergo occlusion in non-muscle cells, any low-level expression of these genes caused by cellular noise is likely to get amplified through a positive feedback loop, which in turn could trigger the erroneous manifestation of muscle phenotype in non-muscle cells. The fact that *Myod1* and *Myf5* are occluded in hLF (and in other non-muscle cell types as shown in what follows) is therefore consistent with the model that the occlusion of key lineage-inappropriate genes serves to restrict cell identity against aberrant transdifferentiation.

Temporal stability of occluded state in fused cells

To investigate how the resetting of gene expression in fused cells is influenced by culture time, we incubated cells for 1, 2, 3, 4, 8 or 16 days after fusion. RT–PCR was used to examine the expression of genes listed in Figure 1A. This showed that the resetting of gene expression occurred mostly within the first 3 days of fusion (data not shown), with expression patterns becoming stabilized after that. Importantly, occluded genes remained silent regardless of post-fusion incubation time (Supplementary Material, Fig. S2), demonstrating the temporal stability of the occluded state in fused cells. This temporal stability is further corroborated by experiments involving the fusion of other cell types (see what follows).

Observed gene occlusion not due to interspecies incompatibility

It is possible that what appears to be the occlusion of hLF genes may actually be the result of interspecies incompatibility—i.e. failure of mouse transcription factors produced from the mSMM genome to recognize the corresponding human *cis*-regulatory sequences in the hLF genome. To address this issue, we fused two cell types that are both of mouse origin but from different strains. One of the two cell types is mSMM, which we have already used, is of C3H strain background. The other is mouse-cloned embryonic fibroblasts (mCEF) of B6 strain background. We exploited sequence polymorphisms between the B6 and C3H mouse strains to determine the origin of transcripts in fused cells.

Among the 24 occluded and 10 transactivated hLF genes, 11 and 3, respectively, were found to be informative in the mCEF–mSMM fusion, meaning that they bear exonic polymorphisms between the two strains based on our resequencing data, and are expressed in mSMM but not mCEF based on RT–PCR data. For each of these genes, RT–PCR primers were designed to flank an inter-strain polymorphic site. The relative abundance of mCEF (B6 strain) versus mSMM (C3H strain) transcripts of the gene in mCEF–mSMM fusion cells was then assessed by sequencing the RT–PCR product. This analysis showed that, of the 11 informative genes occluded in hLF, all but one are also occluded in mCEF based on their exclusive expression from the mSMM

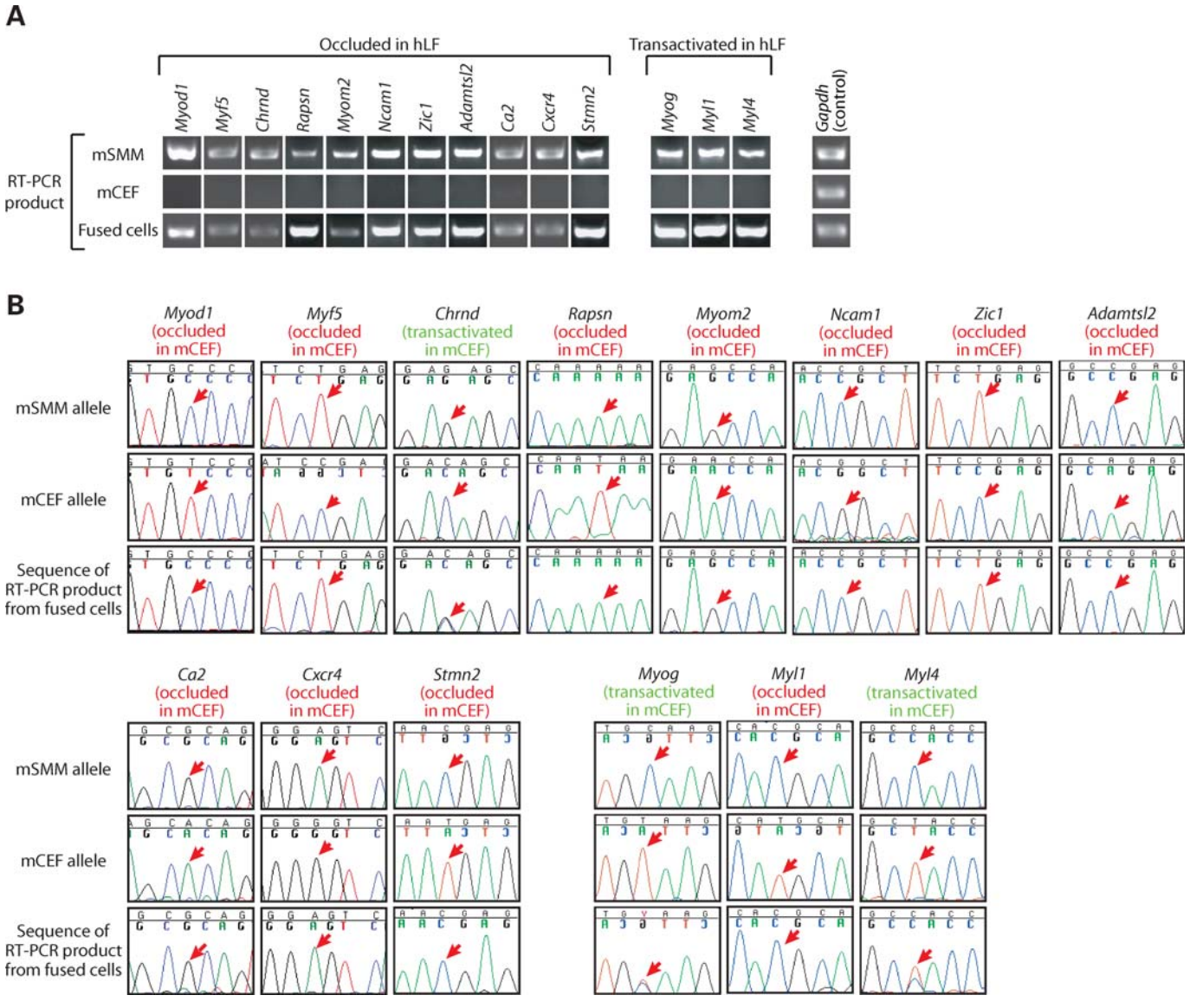

Figure 2. Expression analysis of the mCEF–mSMM fusion by RT–PCR and sequencing on genes found to be occluded or transactivated in hLF. (**A**) RT–PCR performed with primers common to mCEF and mSMM. It showed expression in mSMM and fused cells but not in mCEF. (**B**) Sequencing of RT–PCR products from fused cells (last row of chromatograms). Eleven of the 14 genes are occluded in mCEF, as only the mSMM allele is expressed in fused cells. In contrast, *Chrnd*, *Myog* and *Myl4* are transactivated, as both mCEF and mSMM alleles are expressed. The first two rows of chromatograms are sequences of either mSMM or mCEF alone, showing different alleles between these two cell types. Red arrows in chromatograms indicate sites that are polymorphic between mCEF and mSMM. mCEF, mouse cloned embryonic fibroblasts; mSMM, mouse skeletal muscle myoblasts; hLF, human lung fibroblasts.

allele in fused cells, including the myogenic master triggers *Myod1* and *Myf5* (Fig. 2; data also summarized in Fig. 1B). The single exception is *Chrnd*, which is expressed at roughly equal levels from both mSMM and mCEF alleles, indicating transactivation. Of the three informative genes transactivated in hLF, two were also found to be transactivated in mCEF and one was occluded in mCEF (Fig. 2; also summarized in Fig. 1B). Similar to the hDF–mSMM fusion described earlier, genes found to be occluded in mCEF in the mCEF–mSMM fusion experiment remained silent in fused cells independent of culture time (Supplementary Material, Fig. S3). Thus, among the informative genes, those occluded in hLF are almost all occluded in mCEF and those transactivated in hLF are mostly transactivated in mCEF. These results offer strong evidence that interspecies incompatibility played a negligible role in the identification of occluded genes in the hLF–mSMM fusion, though we cannot rule out the possibility that incompatibility might have affected a small number of genes. The fact that *Chrnd* appears occluded in hLF but transactivated in mCEF suggests the possibility that the observed occlusion of this gene in hLF might be an artifact of interspecies incompatibility in the hLF–mSMM fusion. However, we think that this is unlikely based on data presented in the following section.

Conservation of occluded state across species

Comparison between hLF and mCEF suggests that the set of genes subject to occlusion in a given cell type—fibroblasts in this case—is relatively conserved between divergent species. To further investigate this conservation, we fused chimpanzee dermal fibroblasts (cDF) with human skeletal muscle myoblasts (hSMM) in order to examine whether genes occluded in hLF are also occluded in cDF. The human–chimpanzee genome divergence is about 1/30 of that between human and mouse and is in fact less than the polymorphism levels within many species. Interspecies incompatibility should therefore not be a significant issue in this case.

Of the 24 occluded and 10 transactivated hLF genes, 12 and 8, respectively, were found to be informative in the cDF–hSMM fusion. For these genes, RT–PCR was performed on cDF–hSMM fusion cells using primers common to both species but flanking human–chimpanzee nucleotide substitutions. Sequencing of the RT–PCR products revealed that, of the 12 informative genes occluded in hLF, all are occluded in cDF (Supplementary Material, Fig. S4; summarized in Fig. 1B). The occluded cDF genes include *Chrnd*, which is transactivated in mCEF, suggesting that the occluded status of this gene in hLF is real. Of the eight informative genes transactivated in hLF, six are transactivated in cDF, whereas the other two are occluded in cDF (Supplementary Material, Fig. S4; summarized in Fig. 1B).

Thus, the occluded or transactivated state of genes in hLF is closely recapitulated in both mCEF and cDF, indicating that the set of genes subject to occlusion in a given cell type is strongly conserved across species. Such conservation argues that the occlusion of lineage-inappropriate genes is a highly regulated process with important biological functions.

Effect of DNA synthesis and nuclear merger on the occluded state

In the mCEF–mSMM fusion experiment, we observed that even though the majority of cells were heterokaryons immediately after fusion and FACS purification, most cells became mononucleated after a few days of culture. Furthermore, the average nuclear diameter of these mononucleated cells is ~40% larger than that of either mCEF or mSMM alone (Supplementary Material, Fig. S5A). We suspected that this was due to the formation of a single nucleus from the multiple nuclei in a given fused cell (i.e. nuclear merger). The most likely scenario that multiple nuclei of a heterokaryon could merge is the breakdown and reassembly of the nuclear envelope as the cell undergoes mitosis. For this to occur, cells in the mCEF–mSMM fusion must be capable of DNA synthesis and mitosis. This is consistent with the observation that the mononucleated cells proliferated in number while in culture. By monitoring the incorporation of the thymidine analog 5-bromo-2′-deoxyuridine (BrdU), we confirmed that the majority of fused mCEF–mSMM cells underwent *de novo* DNA synthesis a few days after fusion (Supplementary Material, Figs S5B and C). To further confirm that the single nucleus present in each of the mononucleated cells indeed contains both mCEF and mSMM genomes, we labeled mCEF and mSMM DNA, prior to fusion, with the thymidine analogs 5-chloro-2′-deoxyuridine (CldU) and 5-iodo-2′-deoxyuridine (IdU), respectively. Four days after fusion and FACS purification, cells were co-immunostained for CldU and IdU. For the great majority of mononucleated cells, the nuclei were found to be double-positive for both CldU and IdU, consistent with the merger of the mCEF and mSMM nuclei (Supplementary Material, Fig. S5D).

One complicating factor in identifying occluded genes in fused cells that have undergone mitosis is the possibility of chromosome loss. If some chromosomes are preferentially lost, they would be underrepresented in fused cells and the genes they carry could appear occluded. We addressed this issue by performing PCR on genomic DNA of the fused cells, amplifying across the same polymorphic sites as those interrogated by RT–PCR. Sequencing of PCR products indicated the presence of both alleles at comparable levels for all genes investigated, which are physically scattered across the genome (data not shown). The allele-specific expression seen in Figure 2 is therefore not the result of chromosome loss. These data also argue that the DNA synthesis observed in the fused mCEF–mSMM cells is likely contributed by the replication of both the mCEF and mSMM genomes, because if only one of two genomes has undergone replication, the alleles of the replicating genome should be consistently overrepresented in the genomic PCR product over the alleles of the non-replicating genome, which is not the case.

It is not clear why cells in the mCEF–mSMM fusion can undergo division whereas cells in the other fusion experiments of this study remain largely as mitotically arrested heterokaryons. This notwithstanding, the fact that occluded genes can be uncovered even after heterokaryons have undergone division argues that the occluded state is robust to DNA replication, nuclear merger and changes in the cell cycle state.

Occlusion of muscle-related genes in diverse non-muscle cell types

If the occlusion of muscle-related genes, especially *Myod1* and *Myf5*, indeed serves to safeguard hLF against the accidental activation of myogenic programs, then similar sets of muscle-related genes are likely to be occluded in other cell types of non-myogenic lineages. To test this possibility, we fused mSMM with non-muscle cell types of diverse lineages, and performed RT–PCR to examine whether the 24 genes occluded in hLF are also occluded in these other cell types. The non-muscle cells used included human mesenchymal stem cells (hMSC), human keratinocytes (hKe) and the human cervical cancer cell line Hela. These cells provide a broad representation of both stem cells and differentiated cells, both normal cells and transformed cells and cells derived from different germ layers.

We found that of the nine known muscle-related genes occluded in hLF, the majority are also occluded in all these additional non-muscle cell types, including the myogenic master regulators *Myod1* and *Myf5* (Supplementary Material, Fig. S6). Of the remaining 15 occluded hLF genes not known to be muscle-related, most were either expressed prior to fusion or were transactivated upon fusion in at least one of the non-muscle cell types interrogated. These results support the model that the occlusion of lineage-inappropriate

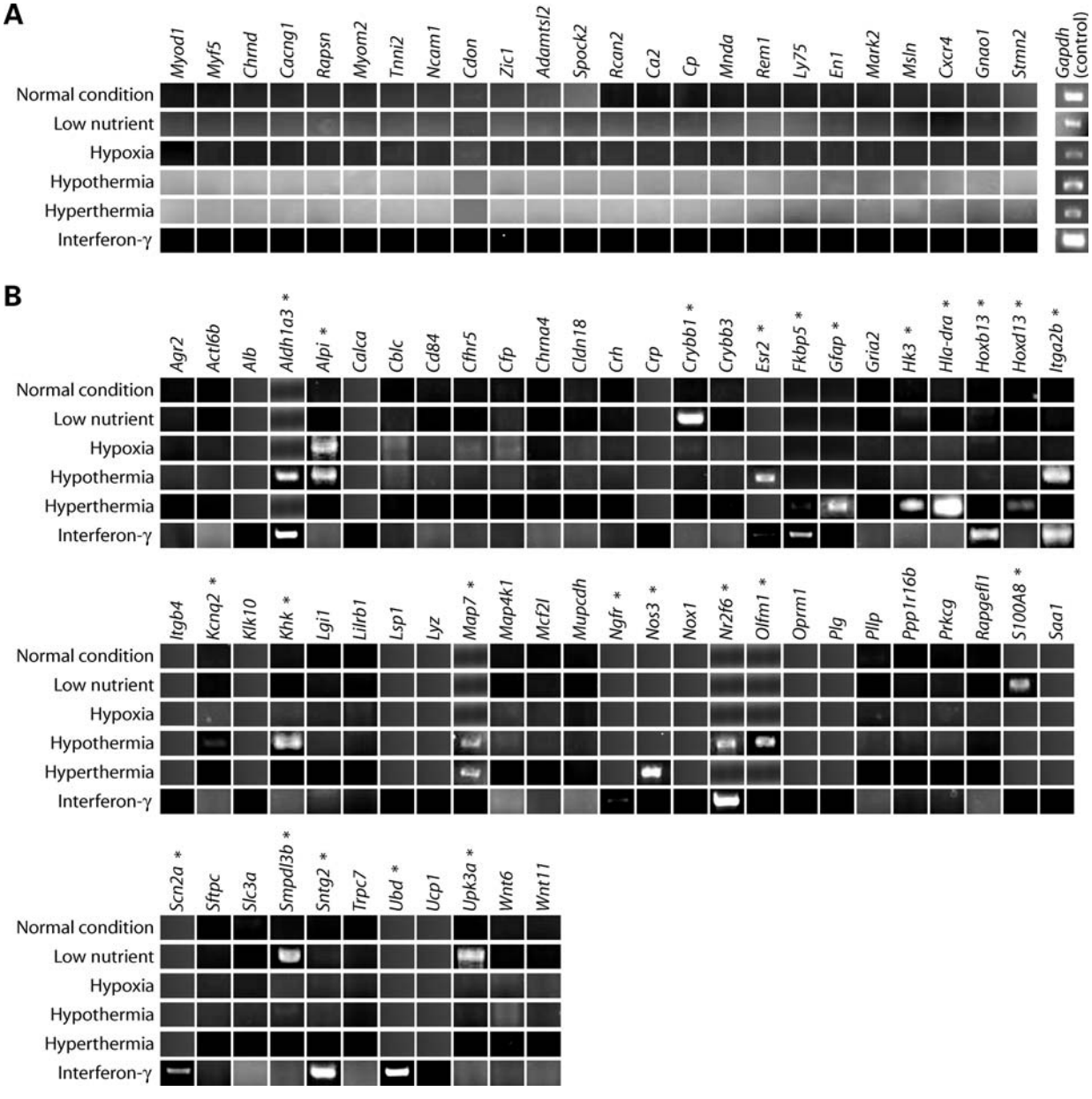

Figure 3. The expression status of occluded genes versus control genes under various culture conditions. Besides the normal culture condition, five additional conditions mimicking physiological variation were used, including low nutrient, hypoxia, hypothermia, hyperthermia and interferon-γ treatment. Control genes were selected on the basis of being silent under the normal condition. (**A**) Stable silencing of occluded genes in hLF under various conditions. (**B**) Activation of some of the control genes in hLF under culture conditions mimicking physiological variation. Genes activated under one or more conditions are indicated by '*'.

genes, especially key master triggers of alternative lineages, contributes to the restriction of cell fate.

Stability of the occluded state under varying physiological conditions

If the occluded state is indeed critical in safeguarding cell identity as we have proposed, then it should be stable under a variety of physiological conditions. To investigate this, we subjected hLF to a variety of culture conditions mimicking various types of physiological stress, including low nutrient, hypoxia, hypothermia and hyperthermia. We also included interferon-γ treatment, which is known to have a dramatic effect on the expression of many genes in a variety of cell types including fibroblasts (30). We then examined the resulting expression patterns of the 24 occluded genes under these culture conditions. All of them remained silent regardless of condition (Fig. 3A). As a control, we identified a set of 61 genes silent in hLF under the normal culture condition based on microarray data and RT–PCR validation. We then examined their expression patterns under the alternative culture conditions. A total of 24 of the 61 genes (39%) became active in at least one of the conditions (Fig. 3B), which is statistically highly distinct from the behavior of zero

activation among the 24 occluded genes ($P < 0.00007$ by Fisher's exact test).

These results demonstrate the extraordinary stability of the occluded state under variable physiological conditions, which stands in stark contrast to the transcriptional lability of other genes in the genome. Researchers have often resorted to genome-wide gene expression patterns (i.e. the transcriptome) as a means of defining cell type identity. However, as our data demonstrate, one cell type has the potential to display considerably different gene expression patterns under different physiological conditions, making the transcriptome too labile to provide a consistent definition of cell type. Our results suggest that genome-wide gene occlusion patterns (i.e. the 'occludome') might provide a much more consistent definition of cell type than the physiologically labile transcriptome (see Discussion).

DISCUSSION

Cell fusion has been used in the past to investigate gene regulation, with most previous studies focusing on the transactivation and extinction of tissue-specific genes in fused cells that indicate the presence of *trans*-acting transcriptional activators or repressors (22,31–33). In this report, we demonstrate an important utility of cell fusion as implemented in the *trans* complementation assay. By fusing disparate cell types and searching for genes differentially expressed between the two genomes of the fused cells, the assay can dissect out the contribution of *cis*-acting mechanisms to gene silencing apart from the contribution of *trans*-acting milieu. Using this assay, we identified a class of genes existing in what we refer to as the occluded state, defined as a state of transcriptional competency whereby a gene remains silent even in the presence of a transcriptionally conducive milieu. We further showed that the occluded state is maintained during cell division and is highly stable under a wide range of physiological conditions.

Monoallelic silencing such as X inactivation and imprinting clearly fits the definition of the occluded state (15–17). Our work demonstrates that biallelic occlusion also occurs as a widespread biological phenomenon, affecting many genes in diverse cell types. Indeed, monoallelic silencing can be viewed as a special case of gene occlusion. It is plausible that biallelic occlusion is the ancestral state that evolved into existence first, with monoallelic silencing evolving subsequently by adopting many of the basic machineries of biallelic occlusion but adding a mechanism for targeting one allele (rather than both alleles) during the silencing process. Biallelic occlusion may be key to defining and restricting the phenotypic identities of cells by stably shutting down lineage-inappropriate genes that might otherwise become active.

Extrapolating from our results, we argue that it may be meaningful to take an 'occludome' perspective of genome regulation—i.e. consider the genome of a cell type as comprising two portions, one being the occluded genes and the other the competent genes (Fig. 4A). Actively expressed genes in a cell type are all competent, but silent genes can be either competent or occluded. This is a different conceptual framework

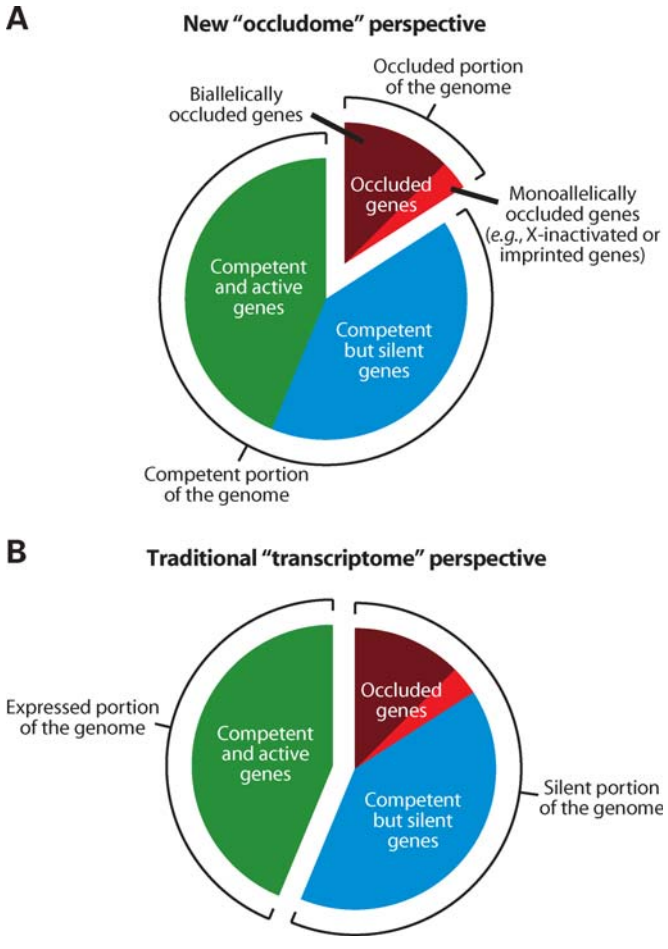

Figure 4. Different perspectives for understanding genome regulation. (**A**) New 'occludome' perspective that considers the genome as consisting of occluded genes and competent genes. (**B**) Traditional 'transcriptome' perspective that views the genome as comprising expressed genes and silent genes. The occludome perspective may provide a molecularly more fundamental and physiologically more stable definition of cell type than the transcriptome.

for understanding genome regulation from the traditional 'transcriptome' perspective whereby genes are considered to be expressed or silent (compare Fig. 4A with B).

We propose that it should be possible to systematically map all the occluded genes in a cell type by fusing it with a wide variety of other cell types that collectively express the entire genome. Such an occludome map might provide a definition of cell type that is physiologically more consistent—and molecularly more fundamental—than the rather labile transcriptome. By comparing occludome maps between cell types of different lineages, between stem cells and differentiated cells of the same lineage, between young and old cells, between normal and pathological cells (such as cancer) and between cells from different species, it might be possible to gain wide-ranging insights into fundamental mechanisms of development, aging, disease processes and evolution. Furthermore, for a given cell type, comparisons could be made between the occludome map and genome-wide maps of chromatin marks such as DNA methylation and histone modifications (34–36). Such comparisons could reveal the biochemical underpinnings of the occluded state and, more importantly,

provide a hitherto unavailable functional readout of the complex chromatin code superimposed on the genetic code.

For operational simplicity, the current study has taken a binary, on/off, view of gene occlusion. However, it is plausible that occlusion can sometimes lead to partial silencing of some genes, in which case a gene may show a quantitative expression difference between the two genomes of fused cells rather than a qualitative on/off difference. In theory, the *trans* complementation assay should be able to reveal both full and partial occlusion as long as a gene displays differential expression between the two genomes of the fused cells (provided that confounding factors such as interspecies incompatibility are ruled out).

The definition of the occluded state requires that a gene is silent (or nearly silent) even in the presence of a transcriptionally conducive milieu. It is important to note, however, that this definition is only in reference to a particular milieu. It may be the case that a gene occluded to one milieu might become active in another milieu. This could happen if transcription factors in the first milieu are blocked by repressive chromatin marks present in certain *cis*-regulatory sequences of a gene, but transcription factors in the second milieu, distinct from the first, are able to drive expression by recognizing a different set of *cis*-regulatory sequences of the gene not affected by repressive chromatin. Alternatively, factors in the second milieu, unlike those in the first, can recognize their target sequences even in the presence of repressive chromatin.

Another important possibility is that some milieus might have the ability to 'deocclude' genes—i.e. erasing the chromatin marks responsible for the occluded state. Such erasure could affect individual genes or the whole genome and could be an active process or a passive one. Reprogramming of somatic cells by nuclear transfer into oocytes or by fusion with embryonic stem cells (ESC) or embryonic germ cells has demonstrated the ability of these cell types to erase most, if not all, of the chromatin marks in somatic cells established during development (37–43). Recent work indicates that such ability may arise from just a few genes whose ectopic expression can reprogram fibroblasts into pluripotent, ESC-like cells called induced pluripotent stem cells (44–50). We hypothesize that shortly after the blastocyst stage (where ESC is derived), cells lose their ability to deocclude the genome, perhaps by occluding the very genes that are responsible for genome-wide deocclusion in the first place. We further hypothesize that the progressive differentiation of cells in subsequent developmental stages is accompanied by the irreversible or nearly irreversible occlusion of an increasing number of genes, with distinct sets of genes becoming occluded in different lineages.

The occlusion of lineage-inappropriate genes could serve to safeguard the phenotypic stability of the myriad cell types in multicellular organisms against noise in both extracellular environment and intracellular regulatory networks. Furthermore, that different cell types are characterized by different occludomes might also explain why the same signaling pathway often triggers the activation of different sets of genes in different cells—a frequent phenomenon during the development of multicellular organisms. The ability of the same transcription factors to play different roles in different cell types allows increased cell type complexity in multicellular organisms without concomitant increases in genome size/complexity. Thus, the evolution of some form of gene occlusion might have been a prerequisite for the evolution of multicellularity.

The occluded state could be quite stable in order to maintain cell identity over the entire ontology of the organism (the germline being an exception where the occluded state is either never fully established for most genes or is erased during gametogenesis). For some genes, the occluded state might be essentially irreversible in somatic cells under normal conditions (as is the case for X-inactivated and imprinted genes). Nevertheless, some occluded genes might become deoccluded in certain somatic cell types by deliberate mechanisms, which could contribute to the dedifferentiation/transdifferentiation of cells during tissue regeneration, especially in species capable of regenerating entire body parts after injury (51). On rare occasions, the competent/occluded status of genes could also change in a stochastic, unregulated manner, which might contribute to aging and disease processes such as cancer. The use of the *trans* complementation assay to systematically identify and characterize occluded genes should therefore have wide-ranging applications in studies of health and disease.

MATERIALS AND METHODS

Cell fusion

mSMM and hLF have been described previously and are known by their common names C2C12 and MRC-5, respectively (21,52). C2C12 (CRL-1772), MRC-5 (CCL-171), Hela (CCL-2), mCEF (TIB-81) and hKe (CRL-2404) were obtained from ATCC; hMSC were derived as described (53) and are available from Cyagen Biosciences; cDF (S006007) were obtained from Coriell Institute for Medical Research; and hSMM (CC-2580T25) were obtained from Cambrex. Cell culture conditions followed published or vendor-supplied protocols.

Neomycin-resistant mSMM cells were generated by transfection with the pEGFP-N1 plasmid (Clontech) and selection in 800 μg/ml G418. EGFP fluorescence varied within this cell population, but was negligible compared with dye fluorescence used for cell sorting. Puromycin-resistant hLF cells were generated using pBabe-puro retroviral vector from Addgene (no. 1764) (54). The vector was transfected into ProPakA.6 packaging cells (ATCC). Forty-eight hours after transfection, the viral supernatant was filtered and added to the cells in the presence of 8 ug/ml polybrene, and 24 h later, cells were selected using 2 μg/ml puromycin.

One day before fusion, cells were labeled with 30 μM CMTMR or 10 μM CMFDA Celltracker dye (Invitrogen) for 30 min at 37°C in culture medium. Subsequently, the cells were incubated in basal medium for 1 h and washed twice with PBS. After staining, mSMM cells were kept in low-serum medium composed of DMEM supplemented with 2% horse serum. Cell fusion was performed with polyethylene glycol (MW 1500) as described (55). Briefly, one of the two cell populations was plated on 10 cm tissue culture dishes and the other cell population was overlaid. After attachment, cells were treated with warm PEG for 1 min and then washed three times with warm basal medium. The fused cells were incubated for

2 h in low-serum medium until cell sorting. Fused cells were purified to >98% purity by FACS, with gating for dual fluorescence. After FACS, purified fused cells were maintained in low-serum medium until RNA extraction. The unfused mSMM used as control in expression studies are kept in low-serum medium for the same period as fused cells. For experiments in which antibiotic-resistant cells were fused, 4–8 μg/ml puromycin and 400–800 μg/ml G418 were added to the medium 1 day after fusion.

Microarray analysis

Total RNA was purified from hLF, mSMM and fused cells using TRIZOL reagent (Invitrogen) according to the vendor's protocol and used for microarray probe synthesis following standard Affymetrix protocols. Double-stranded cDNA samples generated using GeneChip One-Cycle cDNA Synthesis kit, with first strand synthesis using oligo(dT) primers (Affymetrix), were used to synthesize biotin-labeled cRNA using GeneChip IVT Labeling kit (Affymetrix), and then the labeled cRNA samples were fragmented using GeneChip Sample Cleanup Module (Affymetrix). Hybridization, labeling and scanning were all performed by the Protein and Nucleic Acid (PAN) Facility at Stanford University. The labeled cRNA sample from each cell type was hybridized to both of mouse MG U74Av2 and human HG U133A GeneChips (Affymetrix) with replicates to assess gene expression and cross-hybridization between species. Probe-level analyses of the images from scanning of chips were performed using Affymetrix GeneChip Operating Software (GCOS). Similar hybridization procedures were carried out for the hLF-most fusion.

Threshold detection *P*-values were set to assign 'present' ($P < 0.05$), 'marginal' ($0.05 \leq P \leq 0.49$) or 'absent' ($P > 0.49$) decision calls for each gene assigned by MAS 5.0 criteria using GCOS. Filtering gene lists on the basis of absolute decision calls to get a candidate list of occluded genes was performed using GeneSpring (Silicon Genetics). Occluded genes were filtered on the basis of the following criteria: absent calls in all of the replicates with hLF hybridized to human chip, absent calls in all of the replicates with fused cells hybridized to human chip, present or marginal calls in at least one of the replicates with mSMM hybridized to mouse chip and present or marginal calls in at least one of the replicates with fused cells hybridized to mouse chip. Filtering of transactivated genes was performed by (i) comparing genes on the basis of absolute decision calls using GeneSpring with criteria of absent calls in all of the replicates with hLF hybridized to human chip and present or marginal calls in at least one of the replicates with fused cells hybridized to human chip, (ii) selecting genes showing differential expression between the two cell types on the basis of signal intensity after normalization by RMA using RMAexpress (http://rmaexpress.bmbolstad.com) or (iii) comparing the data from the two cell types by model-based expression index analysis using dChip (http://biosun1.harvard.edu/complab/dchip).

RT–PCR and sequencing

RNA (up to 2 μg) was used to generate cDNA using M-MLV reverse transcriptase and random primers (Invitrogen), or using SuperScript III First-Strand Synthesis System with random primers for RT–PCR (Invitrogen) following the vendor's protocol. Semi-quantitative PCR was carried out with variable template concentrations and PCR cycles to obtain linear range amplification of each gene. For the human–chimpanzee fusion experiment, primers were selected by identifying non-polymorphic primer sequences flanking intron-spanning amplicons that contain at least one single-nucleotide substitution between the two species based on genomic sequence alignment. For the mouse–mouse fusion experiment, amplicons containing at least one polymorphism between the two mouse strains were identified by sequencing randomly chosen intron-spanning amplicons. Sequences of PCR primers and detailed conditions for RT–PCR are available upon request. All DNA sequence analysis was performed with the ABI 3730 DNA Analyzer using the ABI BigDye Terminator (Applied Biosystems).

Analysis of DNA synthesis and nuclear merger

For the analysis of nuclear merger, unfused cells were labeled with 10 μM IdU or CldU in the media for 72 h prior to fusion, and fused cells were stained specifically with mouse monoclonal anti-IdU (Becton-Dickinson, 347580; 1:500 dilution) and rat monoclonal anti-CldU antibodies (Accurate, OBT0030; 1:250) on the basis of published protocol (56). These two antibodies do not cross-react when used for double-staining IdU and CldU, but both recognize BrdU. Secondary antibodies were Oregon Green-labeled goat anti-mouse (Invitrogen; 1:1000) and Cy3-labeled mouse anti-rat antibodies (Jackson Immunoresearch; 1:300). For the analysis of DNA synthesis, BrdU was administered at 10 μM in the media immediately following cell fusion for 72 h, and the cells stained with anti-BrdU antibody (Accurate, OBT0030; 1:250) at a later time point. Because the incorporation of halogenated nucleotides into DNA could affect gene expression, the fusion experiment involving labeling with halogenated nucleotides is done separately from the fusion experiment for ascertaining the expression status of genes.

Analysis of gene expression under physiological alterations

Cells were cultured under either the normal condition (10% fetal calf serum at 37°C) or one of the conditions mimicking physiological alterations, including low nutrient (0.1% serum), hypoxia (380 μM of the hypoxia mimetic deferoxamine), hypothermia (33°C), hyperthermia (41°C) and interferon-γ treatment (100 ng/ml; Cell Sciences). Cells were maintained under each condition for 3 days, followed by RT–PCR analysis of selected genes as described earlier.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Ryan Duggan, David Leclerc, Michael Olson, Jaejung Kim and Ingrid Kilgore for technical assistance.

Conflict of Interest statement. None declared.

FUNDING

This work was partly supported by National Institutes of Health grants F32HL922792 (to J.G.) and F32GM075503 (to G.E.S.). Funding to pay the Open Access charge was provided by Howard Hughes Medical Institute.

REFERENCES

1. Waddington, C.H. (1966) *Principles of Development and Differentiation*. Macmillan, New York.
2. Caplan, A.I. and Ordahl, C.P. (1978) Irreversible gene repression model for control of development. *Science*, **201**, 120–130.
3. Fisher, A.G. and Merkenschlager, M. (2002) Gene silencing, cell fate and nuclear organisation. *Curr. Opin. Genet. Dev.*, **12**, 193–197.
4. Macaluso, M. and Giordano, A. (2004) How does DNA methylation mark the fate of cells? *Tumori*, **90**, 367–372.
5. Sparmann, A. and van Lohuizen, M. (2006) Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer*, **6**, 846–856.
6. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
7. Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, **33** (Suppl.), 245–254.
8. Vermaak, D., Ahmad, K. and Henikoff, S. (2003) Maintenance of chromatin states: an open-and-shut case. *Curr. Opin. Cell Biol.*, **15**, 266–274.
9. Goldberg, A.D., Allis, C.D. and Bernstein, E. (2007) Epigenetics: a landscape takes shape. *Cell*, **128**, 635–638.
10. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
11. Li, B., Carey, M. and Workman, J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707–719.
12. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. and Cavalli, G. (2007) Genome regulation by polycomb and trithorax proteins. *Cell*, **128**, 735–745.
13. Surani, M.A., Hayashi, K. and Hajkova, P. (2007) Genetic and epigenetic regulators of pluripotency. *Cell*, **128**, 747–762.
14. Bird, A. (2007) Perceptions of epigenetics. *Nature*, **447**, 396–398.
15. Bartolomei, M.S. and Tilghman, S.M. (1997) Genomic imprinting in mammals. *Annu. Rev. Genet.*, **31**, 493–525.
16. Goldmit, M. and Bergman, Y. (2004) Monoallelic gene expression: a repertoire of recurrent themes. *Immunol. Rev.*, **200**, 197–214.
17. Valley, C.M. and Willard, H.F. (2006) Genomic and epigenomic approaches to the study of X chromosome inactivation. *Curr. Opin. Genet. Dev.*, **16**, 240–245.
18. Barrera, L.O. and Ren, B. (2006) The transcriptional regulatory code of eukaryotic cells—insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr. Opin. Cell Biol.*, **18**, 291–298.
19. Maziarz, M., Chung, C., Drucker, D.J. and Emili, A. (2005) Integrating global proteomic and genomic expression profiles generated from islet alpha cells: opportunities and challenges to deriving reliable biological inferences. *Mol. Cell. Proteomics*, **4**, 458–474.
20. Chiu, C.P. and Blau, H.M. (1984) Reprogramming cell differentiation in the absence of DNA synthesis. *Cell*, **37**, 879–887.
21. Blau, H.M., Pavlath, G.K., Hardeman, E.C., Chiu, C.P., Silberstein, L., Webster, S.G., Miller, S.C. and Webster, C. (1985) Plasticity of the differentiated state. *Science*, **230**, 758–766.
22. Boshart, M., Nitsch, D. and Schutz, G. (1993) Extinction of gene expression in somatic cell hybrids—a reflection of important regulatory mechanisms? *Trends Genet.*, **9**, 240–245.
23. Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, **51**, 987–1000.
24. Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E. and Arnold, H.H. (1989) A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. *EMBO J.*, **8**, 701–709.
25. Wright, W.E., Sassoon, D.A. and Lin, V.K. (1989) Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. *Cell*, **56**, 607–617.
26. Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H.H. (1990) Myf-6, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *EMBO J.*, **9**, 821–831.
27. Miner, J.H. and Wold, B. (1990) Herculin, a fourth member of the MyoD family of myogenic regulatory genes. *Proc. Natl Acad. Sci. USA*, **87**, 1089–1093.
28. Thayer, M.J., Tapscott, S.J., Davis, R.L., Wright, W.E., Lassar, A.B. and Weintraub, H. (1989) Positive autoregulation of the myogenic determination gene MyoD1. *Cell*, **58**, 241–248.
29. Tapscott, S.J. (2005) The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development*, **132**, 2685–2695.
30. Der, S.D., Zhou, A., Williams, B.R. and Silverman, R.H. (1998) Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, **95**, 15623–15628.
31. Davidson, R.L. (1974) Gene expression in somatic cell hybrids. *Annu. Rev. Genet.*, **8**, 195–218.
32. Blau, H.M. (1989) How fixed is the differentiated state? Lessons from heterokaryons. *Trends Genet.*, **5**, 268–272.
33. Gourdeau, H. and Fournier, R.E. (1990) Genetic analysis of mammalian cell differentiation. *Annu. Rev. Cell Biol.*, **6**, 69–94.
34. van Steensel, B. and Henikoff, S. (2003) Epigenomic profiling using microarrays. *Biotechniques*, **35**, 346–350, 352–344, 356–347.
35. Bernstein, B.E., Meissner, A. and Lander, E.S. (2007) The mammalian epigenome. *Cell*, **128**, 669–681.
36. Jones, P.A. and Baylin, S.B. (2007) The epigenomics of cancer. *Cell*, **128**, 683–692.
37. Gurdon, J.B. (1962) The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *J. Embryol. Exp. Morphol.*, **10**, 622–640.
38. Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J. and Campbell, K.H. (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature*, **385**, 810–813.
39. Hochedlinger, K. and Jaenisch, R. (2006) Nuclear reprogramming and pluripotency. *Nature*, **441**, 1061–1067.
40. Tada, M., Tada, T., Lefebvre, L., Barton, S.C. and Surani, M.A. (1997) Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J.*, **16**, 6510–6520.
41. Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. and Tada, T. (2001) Nuclear reprogramming of somatic cells by *in vitro* hybridization with ES cells. *Curr. Biol.*, **11**, 1553–1558.
42. Do, J.T. and Scholer, H.R. (2004) Nuclei of embryonic stem cells reprogram somatic cells. *Stem Cells*, **22**, 941–949.
43. Cowan, C.A., Atienza, J., Melton, D.A. and Eggan, K. (2005) Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science*, **309**, 1369–1373.
44. Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**, 663–676.
45. Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R. *et al.* (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell*, **1**, 55–70.
46. Okita, K., Ichisaka, T. and Yamanaka, S. (2007) Generation of germline-competent induced pluripotent stem cells. *Nature*, **448**, 313–317.
47. Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E. and Jaenisch, R. (2007) *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*, **448**, 318–324.
48. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861–872.
49. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R. *et al.* (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science*, **318**, 1917–1920.

50. Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S. (2007) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.*, **26**, 101–106.
51. Sanchez Alvarado, A. and Tsonis, P.A. (2006) Bridging the regeneration gap: genetic insights from diverse animal models. *Nat. Rev. Genet.*, **7**, 873–884.
52. Yaffe, D. and Saxel, O. (1977) Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*, **270**, 725–727.
53. Zhang, A.X., Yu, W.H., Ma, B.F., Yu, X.B., Mao, F.F., Liu, W., Zhang, J.Q., Zhang, X.M., Li, S.N., Li, M.T. *et al.* (2007) Proteomic identification of differently expressed proteins responsible for osteoblast differentiation from human mesenchymal stem cells. *Mol. Cell. Biochem.*, **304**, 167–179.
54. Morgenstern, J.P. and Land, H. (1990) Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res.*, **18**, 3587–3596.
55. Davidson, R.L., O'Malley, K.A. and Wheeler, T.B. (1976) Polyethylene glycol-induced mammalian cell hybridization: effect of polyethylene glycol molecular weight and concentration. *Somatic Cell Genet.*, **2**, 271–280.
56. Vega, C.J. and Peterson, D.A. (2005) Stem cell proliferative history in tissue revealed by temporal halogenated thymidine analog discrimination. *Nat. Methods*, **2**, 167–169.