2008-09-24

Gene associations: true romance or chance meeting in a nuclear neighborhood

Jeanne B. Lawrence
University of Massachusetts Medical School

Christine Moulton Clemson
University of Massachusetts Medical School

Published September 22, 2008

Gene associations: true romance or chance meeting in a nuclear neighborhood?

Jeanne B. Lawrence and Christine M. Clemson

Department of Cell Biology, University of Massachusetts Medical Center, Worcester, MA 01655

Many recent studies have raised interest in the nuclear associations of coregulated genes from different chromosomes, often evoking interpretations of gene–gene interactions, communication, and even "romance." However, in some cases, the associations may be indirect and infrequent and may reflect the segregation of active and inactive genes into different nuclear compartments. The study by Brown et al. (see p. 1083 of this issue) reports that the apparent association of erythroid genes is not a direct interaction nor colocalization to one tiny transcription factory but arises as a result of the known clustering of many active genes with larger splicing factor–rich speckles (a.k.a., SC35-defined domains). This clustering appears largely stochastic but is impacted by the chromosomal neighborhood of the gene as well as its transcriptional status. The study adds a new twist by examining the same gene in a foreign chromosomal context, providing evidence that this impacts a gene's propensity to form gene–domain (or apparent gene–gene) associations within nuclei.

Interpreting gene associations in nuclei by FISH

Gene pairing in mammalian nuclei has been implicated in processes ranging from transient "kissing" of alleles during X inactivation or imprinted gene regulation to the colocalization of unlinked genes that are tightly coregulated during differentiation (for review see de Laat and Grosveld, 2007; Brown et al., 2008). Evidence for interallelic or interchromosomal gene associations in mammalian nuclei is based on two types of analyses: simultaneous visualization of loci by FISH or the use of 3C/4C technology to identify DNA sequences close enough for paraformaldehyde cross-linking and subsequent ligation (Dekker, 2006; for review see de Laat, 2007). In situ analysis is thought to be important for validation of the more molecular-level associations identified by 3C/4C, and 3C can similarly be used to validate apparent in situ interactions. However, an oft used definition of association is a separation of up to 0.5–1.0 µm between gene signals by FISH, allowing a cytological separation inconsistent with a molecular scale interaction. In some studies, signals that are closely juxtaposed are used to indicate colocalization, but given the limit of resolution of light microscopy (0.1–0.2 µm), any separation (lack of overlap) in two signals suggests a significant molecular-scale disconnect. For example, two 10-kb probes separated by 100–200 kb on a chromosome can only be separately resolved in a fraction of nuclei, even though these sequences may still be distant and noninteracting on a molecular scale (Lawrence et al., 1990). Brown et al. (2008) recognize that a direct physical interaction between two DNA sequences would generate completely overlapping FISH signals, which they essentially do not see (Brown et al., 2006, 2008). Thus, they examine whether the increased proximity of active genes may be accounted for by their association with other nuclear structures.

Globin gene associations may reflect the interaction of active erythroid genes with relatively large SC35-rich speckles

A prior study reported that as globin genes are activated in erythroid cells, they become colocalized with each other within a single transcription factory (Osborne et al., 2004). Transcription factories, initially defined by bromo-UTP (BrUTP) labeling for RNA synthesis, appear as ∼2,000 closely spaced punctuate spots that by electron microscopy are ∼40 nm (Pombo et al., 1999) in most cell types, which is as small or smaller than a single gene FISH signal. Although Brown et al. (2008) confirm that the erythroid genes are more associated by FISH in erythroid cells versus lymphocytes, they argue that the numbers do not add up: signals very rarely colocalize, and even associated signals are on average still a whopping 0.5 µm apart, ∼10 times the size of transcription factories in human erythroid cells. Brown et al. (2008) further suggest that the polymerase II antibody used in this prior analysis does not detect the transcription factories defined by BrUTP incorporation. Rather, they conclude that what brings two active loci in proximity is their mutual association with nuclear speckles, also known as SC35 domains. These large (0.5–2.0 µm) structures enriched in numerous splicing and RNA metabolic factors are often delineated by staining to spliceosome assembly factor SC35, which typically

Correspondence to Jeanne B. Lawrence: Jeanne.Lawrence@umassmed.edu

The Rockefeller University Press $30.00
J. Cell Biol. Vol. 182 No. 6 1035–1038
www.jcb.org/cgi/doi/10.1083/jcb.200808121

Published September 22, 2008

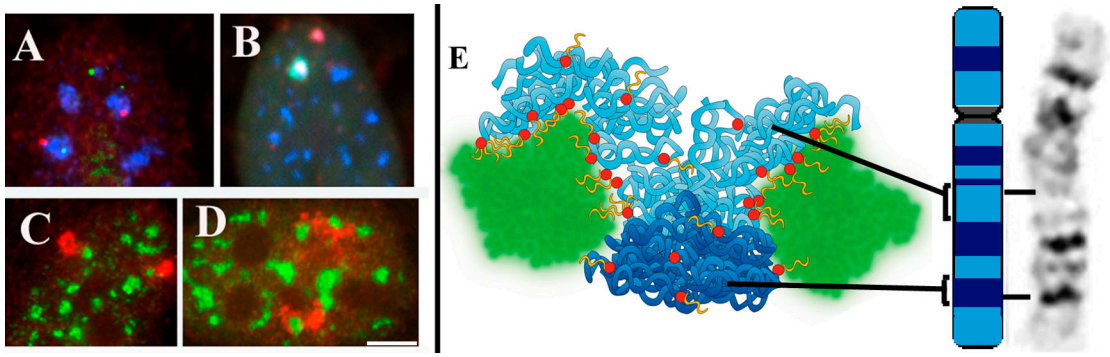

Figure 1. **Localization of active genes or chromosome regions with splicing factor (SC35)–rich domains impacts their overall nuclear organization.** (A) *COL1A1* and *COL1A2* genes on separate chromosomes can associate with a common SC35 domain, increasing their proximity to each other. Human diploid fibroblasts hybridized with genomic probes to *COL1A1* (red) and *COL1A2* (green) and stained for SC35 (blue). One allele of each gene is positioned at the periphery of the same SC35 domain in the cell shown. Although this brings these loci closer, they remain separate. (B) *COL1A1* and *COL1A2* RNA foci can be completely coincident within an SC35 domain. Transcripts from the *COL1A1* (green) and *COL1A2* (red) genes enter the SC35 domain and can completely overlap within this common SC35 domain (blue). Although RNA foci completely overlap, genes remain separate. Overlap between the three colors appears white. (C–E) Clustering of genes in bands on chromosomes may relate to their clustering with nuclear substructures. (C) The G-band probe for 3p14 (red) shows two relatively compact bands minimally contacting SC35 domains (green). (D) The probe for a similarly sized segment of R-band 17q21 DNA (red) shows two signals that are much more highly extended and show more regions intimately contacting SC35 domains. (E) Model for the differential organization of R- and G-band DNA with respect to SC35-rich domains and euchromatic neighborhoods. Multiple genes cluster at the periphery of a single, very large accumulation (domain) of mRNA metabolic factors, forming a more euchromatic neighborhood around the domain. R-band DNA (light blue), which is somewhat more gene rich, is more intimately associated with these SC35 domains than more tightly packaged G-band DNA (dark blue). (right) Ideogram and actual appearance of cytogenetic bands and sub-bands are shown for chromosome 17. This figure is modified from Shopland et al. (2003). Bar, 5 µm.

identifies ~15–40 SC35 domains per nucleus (for review see Lamond and Spector, 2003; Hall et al., 2006). Although the function of these speckles is debated, the Brown et al. (2008) study adds to the evidence that speckles are associated with many highly active genes.

If globin genes interact with a larger SC35 speckle rather than smaller and more numerous transcription factories, as Brown et al. (2008) conclude, this is an important distinction. As shown in Fig. 1 A, two genes that associate with the same SC35 domain can be proximal to each other but still often quite far apart. This does not imply that the genes are interacting with each other or need to be near one another to be coregulated, as is more strongly suggested from colocalization to a transcription factory. Genes typically position at the outer edge of the SC35 domain, whereas the transcripts (from genes associated with that domain) often enter the domain and thus can transiently comingle there (Shopland et al., 2003; Hall et al., 2006). As shown here for COL1 and COL2 (Fig. 1, A and B), the two RNA foci can overlap within the same SC35 domain, whereas the genes are still separable. Thus, experiments that use RNA foci to mark specific genes (Osborne et al., 2007) may overestimate the frequency of truly coincident genes.

Is speckle choice random?

Even by the liberal 1-µm definition of DNA–DNA association, most nuclei still do not show associations between any two given loci. Association rates are typically around 10%, with the highest in Brown et al. (2008) approaching 30%. Importantly, Brown et al. (2008) find that the subset of gene pairs that are in proximity are consistently contacting an SC35-defined speckle, indicating that this is what drives the "pairing." The clustering of active erythroid genes with SC35 domains fits with several prior studies, which showed that localization at the immediate edge of these domains is a common but locus-specific organization that closely correlates with the activity of some genes (Jolly et al., 1999; Smith et al., 1999; Hall et al., 2006; Brown et al., 2006). Throughout these studies, gene–domain associations have been more tightly defined (no visible separation) by fluorescence microscopy and commonly occur at much higher rates (~90–100%) than gene–gene associations; in fact, gene–domain association rates of 5–20% were generally considered random.

If two active genes associate with speckles with high frequency, why are they not more frequently associated with each other? Insight into this was provided for the tightly coregulated COL1A1 and COL1A2 genes system (Shopland et al., 2003). Both genes associate with an SC35 domain almost 100% of the time, yet they were at the same SC35 speckles in only 7–13% of nuclei. Thus, rather than conclude that the genes had an affinity for one another, this nonrandom proximity was interpreted to be a biproduct of the specific positioning of both genes at SC35 domains. Because the choice of the domain appeared essentially random (in cells with 15–20 domains), these two tightly coregulated genes did not interact and communicate with each other but were expressed in a common subnuclear compartment. Although not precluding that instances of more specific gene pairing may occur in mammalian cells, this demonstrates that apparent gene–gene associations can arise as a result of broader principles of gene organization in nuclei.

3C analysis of gene associations: are weak interactions mediated by a third party?

Although the study by Brown et al. (2008) does not itself repeat 3C analysis of the erythroid genes, α- and β-globin genes have been perhaps the most studied genes by this method (Dekker, 2006; for review see de Laat, 2007). This approach has provided important evidence of direct interlocus interactions; for example, through chromatin looping of the locus control region and promoter in the mouse β-globin locus (Vakoc et al., 2005) or physical

Published September 22, 2008

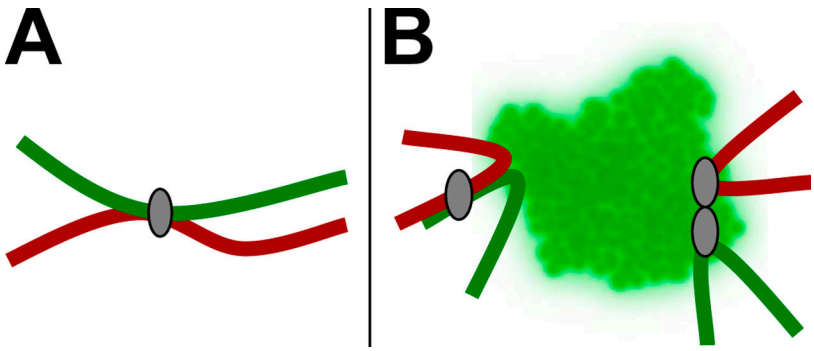

Figure 2. **Alternative explanations for interchromosomal gene associations seen by 3C analysis.** (A) A direct molecular interaction of two gene loci occurs such that the formaldehyde cross-links form (gray) via their chromatin proteins. (B) Two DNA segments that localize to a common nuclear structure will infrequently be near enough to one another to be cross-linked via their chromatin proteins (left) or to be indirectly connected via mutual cross-links to the larger aggregation of splicing factors or other protein accumulations (right). The latter may involve multiple cross-links, but subsequent ligation and PCR amplification would occur in the absence of a direct in vivo molecular interaction.

interaction of elements with the Ig heavy chain gene cluster (Ju et al., 2007). Using 3C techniques, interchromosomal DNA associations for the α- and β-globin genes and a number of other loci have been reported (Osborne et al., 2004; for review see de Laat and Grosvfeld, 2007). However, as recently discussed (Dekker, 2006), there may be a host of weaker interactions detected by 3C/4C that arise as a result of very infrequent and possibly quite indirect interactions. The specificity of these interactions may in fact be quite limited: genes that are active may be more likely to become cross-linked with each other because of their interaction with a third party, such as a common nuclear domain (Fig. 2). Similarly, heterochromatic genes would have chance 3C interactions as a result of their segregation in a distinct nuclear compartment, or "neighborhood." Thus, in the case of weaker 3C interactions, it is all the more necessary to have strong FISH data to corroborate that these associations are meaningful.

In some cases, very low frequency associations could reflect transient one-time regulatory events, such as the initial choice of which X chromosome to inactivate or which imprinted allele will be expressed. It is harder to see how this would relate to the coregulation of two or more genes that are both stably expressed (or repressed) during development because, even using relatively loose criteria by FISH, only a subset of these expressing genes are typically paired with one another.

The influence of the chromosomal context

Brown et al. (2008) report that all five erythroid genes examined showed increased proximity in the active state, but the extent of this varied with the gene and, to some extent, the species. To investigate whether this is influenced by chromosomal context, the authors examine the 2-Mb chromatin regions around each locus for gene density, CpG islands, and the degree of distension via FISH. The two genes that showed the greatest propensity for interchromosomal gene associations were also in chromosomal regions that show the greatest extension within nuclei. Although the sample size is small, their evidence is bolstered by another approach to examine cells in which the human α-globin locus was placed into the mouse α-globin chromosomal site, which shows less frequent interactions than its human counterpart. The human α gene associated with a frequency similar to the mouse locale, providing further evidence that the broader chromosomal context can impact such interactions.

Genomic scale studies examining interchromosomal interactions by 4C have also begun to identify more interacting regions of chromosomes. For example, organizational differences between relatively large chromosomal segments is suggested by 4C studies (Simonis et al., 2006; for review see de Laat, 2007). Some chromosomal regions are more apt to show a 4C signal above background, and this can vary with cell type. More active regions cross-link more with active regions, and silenced regions cross-link more with silent regions. However, a question remains as to whether this reflects sequence-specific interactions or rather the general principles of nuclear organization in which genes segregate into larger euchromatic versus heterochromatic compartments.

Prior studies showed that a consistent association of genes with SC35 domains was not a property of all active genes but was locus specific and may relate to the properties of the gene or its RNA (Smith et al., 1999) as well as chromosomal neighborhood or band (Shopland et al., 2003). In fact, differences in association levels between regions of a chromosome reported by both Brown et al. (2008) and Simonis et al. (2006) may well relate to a long-known feature of chromosome biology: cytogenetic bands. The occurrence of chromosome bands itself demonstrates that there is a higher order organization of DNA in large (~2–10 Mb) blocks (Fig. 1 E). A prior study found that R bands were more extended and intimately associated with SC35-rich speckles than G bands (Fig. 1, C and D; Shopland et al., 2003) and suggested that the organization manifest in chromosome bands may have evolved in part to accommodate the clustered organization of genes into nuclear neighborhoods. Accordingly, Brown et al. (2008) found that the most interactive gene, SLC4A1, maps to 17q21, an R band that Shopland et al. (2003) showed was especially distended and associated with SC35 speckles.

In fact, constraints imposed by chromosomal context likely impede the inherent propensity of active sequences to cluster, as illustrated by the fact that transfected sequences are not transcribed throughout the nucleus but are found in large nuclear clusters. An interesting recent study by Xu and Cook (2008) indicates that plasmids containing similar promoter types cluster together, forming foci that largely overlap with polymerase II as well as SC35.

Relation of gene organization to expression

A key question in any study of gene organization is how it relates to gene expression and regulation. Although not the main focus of Brown et al. (2008), their findings are consistent with a host of evidence that transcriptional activity correlates with the

Published September 22, 2008

increased proximity of coregulated genes, gene positioning at the edge of SC35 speckles, or localization with transcription factories. However, this in itself does not demonstrate that the organization is required for activity. The available evidence has long indicated that expression of a gene does not require its association with an SC35-rich speckle, and Brown et al. (2008) make the case that the α- and β-globin genes still produce transcription foci at high frequency, irrespective of their different frequencies of association or their different chromosomal context. In contrast, other studies conclude that the movement into and out of preexisting transcription factories is required for gene expression (Osborne et al., 2004, 2007). Although many factors will impact the level of expression from a gene, it is likely that nuclear organization and chromosomal context both have a significant impact. Although RNA FISH can determine how many loci are associated with visible RNA, simply counting RNA foci associated with genes will not actually determine whether the level of expression (and posttranscriptional processing) changes. Therefore, a more challenging question for future studies is to determine whether the level of expression is indeed influenced by nuclear and chromosomal organization. Basic kinetic principles would suggest that the immediate access to high concentrations of numerous RNA metabolic factors would be expected to increase the efficiency of the enzymatic processes required.

Irrespective of whether genes interact with each other, an SC35-enriched speckle, or a transcription factory, the nonrandom organization of genes in nuclei is interesting and likely important in gene regulation. Although there is still much to learn from the detailed analysis of endogenous genes in their native context, Brown et al. (2008) illustrates an initial foray into what will be an important approach in the future: "locus-switching" experiments to test predictions about the impact on nuclear organization as well as expression of the chromosomal neighborhood.

Submitted: 22 August 2008
Accepted: 29 August 2008

References

Brown, J.M., J. Leach, J.E. Reittie, A. Atzberger, J. Lee-Prudhoe, W.G. Wood, D.R. Higgs, F.J. Iborra, and V.J. Buckle. 2006. Coregulated human globin genes are frequently in spatial proximity when active. *J. Cell Biol.* 172:177–187.

Brown, J.M., J. Green, R. Pires das Neves, H.A.C. Wallace, A.J.H. Smith, J. Hughes, N. Gray, S. Taylor, W.G. Wood, D.R. Higgs, F.J. Iborra, and V.J. Buckle. 2008. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* 182:1083–1097.

de Laat, W. 2007. Long-range DNA contacts: romance in the nucleus? *Curr. Opin. Cell Biol.* 19:317–320.

de Laat, W., and F. Grosveld. 2007. Inter-chromosomal gene regulation in the mammalian cell nucleus. *Curr. Opin. Genet. Dev.* 17:456–464.

Dekker, J. 2006. The three 'C's of chromosome conformation capture: controls, controls, controls. *Nat. Methods.* 3:17–21.

Hall, L.L., K.P. Smith, M. Byron, and J.B. Lawrence. 2006. Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288:664–675.

Jolly, C., C. Vourc'h, M. Robert-Nicoud, and R.I. Morimoto. 1999. Intron-independent association of splicing factors with active genes. *J. Cell Biol.* 145:1133–1143.

Ju, Z., S.A. Volpi, R. Hassan, N. Martinez, S.L. Giannini, T. Gold, and B.K. Birshtein. 2007. Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3′ regulatory region. *J. Biol. Chem.* 282:35169–35178.

Lamond, A.I., and D.L. Spector. 2003. Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4:605–612.

Lawrence, J.B., R.H. Singer, and J.A. McNeil. 1990. Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science.* 249:928–932.

Osborne, C.S., L. Chakalova, K.E. Brown, D. Carter, A. Horton, E. Debrand, B. Goyenechea, J.A. Mitchell, S. Lopes, W. Reik, and P. Fraser. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36:1065–1071.

Osborne, C.S., L. Chakalova, J.A. Mitchell, A. Horton, A.L. Wood, D.J. Bolland, A.E. Corcoran, and P. Fraser. 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* 5:e192.

Pombo, A., M. Hollinshead, and P.R. Cook. 1999. Bridging the resolution gap: imaging the same transcription factories in cryosections by light and electron microscopy. *J. Histochem. Cytochem.* 47:471–480.

Shopland, L.S., C.V. Johnson, M. Byron, J. McNeil, and J.B. Lawrence. 2003. Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* 162:981–990.

Simonis, M., P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Wit, B. van Steensel, and W. de Laat. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38:1348–1354.

Smith, K.P., P.T. Moen, K.L. Wydner, J.R. Coleman, and J.B. Lawrence. 1999. Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J. Cell Biol.* 144:617–629.

Vakoc, C.R., D.L. Letting, N. Gheldof, T. Sawado, M.A. Bender, M. Groudine, M.J. Weiss, J. Dekker, and G.A. Blobel. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell.* 17:453–462.

Xu, M., and P.R. Cook. 2008. Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* 181:615–623.