Wen *et al.* reply:

Filion and van Steensel claim that the differentiation-related large organized chromatin K9 modifications (LOCKs) we reported[1] are not supported by our microarray data. We disagree, but also note that our conclusions regarding LOCKs were not based on array data alone but also on the many validations and functional experiments described in the paper, including real-time PCR validation, conservation of LOCKs, and genetic knockout of the histone methyltransferase G9a in ES cells and this knockout's influence on gene expression. Interestingly, these functional data are not challenged by Filion and van Steensel, nor do they question the existence of LOCKs or the tissue specificity of LOCKs. **Figure 1a** and **Supplementary Figure 1** show replicate data, available on the web page cited in the paper (http://rafalab.jhsph.edu/k9LOCKs/), superimposed on the genome-wide data from the paper. The replicate data include two examples of undifferentiated ES cell lines cultured separately (biological replicates), as well as two differentiated cell types (differentiated ES cells and liver cells). In all cases where differences between undifferentiated and differentiated ES cells were described in the paper, the same differences are seen even more dramatically in the replicate experiments. Notably, the replicates are quite consistent for undifferentiated ES cells, and their signals were dwarfed by the relative signal in differentiated cells. The probability of observing by chance these same differences, with the same termini, is extremely low. We also performed quantitative real-time PCR validation of the LOCKs comparing chromatin immunoprecipitated DNA from undifferentiated and differentiated ES cells; this showed unequivocally that the LOCKs are differentiation specific (**Fig. 1b**).

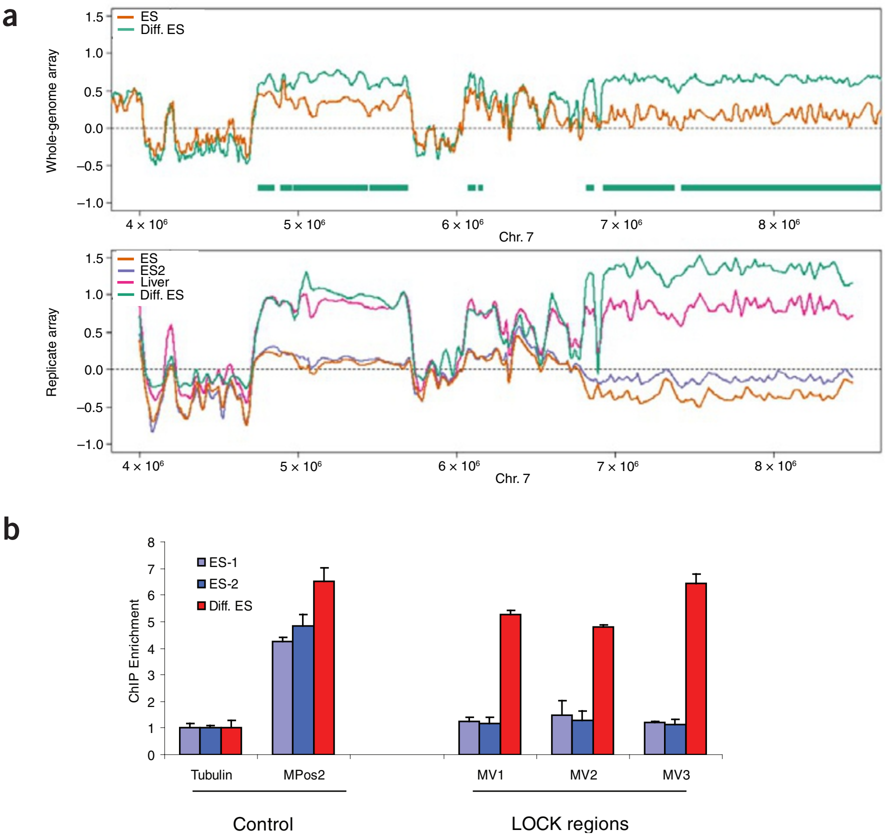

Figure 1 Replicate data support differences of LOCKs between undifferentiated and differentiated ES cells. (**a**) Comparison of genome-wide array and replicate array. Upper plot shows smoothed curves of undifferentiated ES (orange) and differentiated ES (green) and regions defined as LOCKs in the paper. Bottom plot shows curves of two independent ES cell cultures (blue and orange), differentiated ES cells (green) and liver (pink) in the same region. The custom array is as described in the paper. (**b**) qPCR validation of differentiated ES cell–specific LOCKs. The *y* axis shows ratios of ChIP/Input normalized to tubulin. Tubulin and MPos2 are negative and positive controls, respectively. MV1–MV3 are three regions in differentiated ES cell–specific LOCKs (MV1 and MV2 from LOCK at chr. 7: 7416826-9017990; MV3 from LOCK at chr. 1: 108835775-109160948). At least three replicate experiments were performed in each case. Primer sequences are available on request.

Regarding microarrays, Filion and van Steensel claim that variations between undifferentiated and differentiated ES cells are due to sample labeling or hybridization conditions. Our extensive experience with microarray data is that variation due to hybridization and labeling can be controlled by appropriately normalizing the data. Our group included expert statisticians and spent a great deal of time and thought on the statistical analysis. Our method was based on a completely data-driven procedure, as described in the Supplementary Methods section of the paper. Notice that the procedure and cutoff we used in the paper worked remarkably well at detecting locations with differential LOCKs between samples as confirmed by (i) RT-PCR for gene expression, (ii) association with gene expression in tissues and (iii) concordance with replicate experiments (ref. 1, **Fig. 1** and **Supplementary Fig. 1**). Comparing various tissue-specific LOCKs using this same procedure revealed an extremely strong relationship between our statistical criteria of LOCKs and domain-specific gene silencing. We did consider other percentiles (cutoffs), and the conclusion that differentiated cells had more LOCKs did not change (**Supplementary Fig. 2**).

Note that the use of a two-state hidden Markov model (HMM) to assess specific microarray signals is not always appropriate. In our original study[1], we specifically said that the LOCKs detected in the genome-wide arrays are not necessarily absent in undifferentiated ES cells but may be minimally present compared to differentiated ES cells, as shown in our original Figure 3a. It is well known that ES cell cultures are usually contaminated with differentiated cells (typically 10% and often substantially more) even when clones are chosen for an apparently undifferentiated morphology[2–4]—which, in fact, as we reported, we did not do. Kalmar *et al.* have now proven that ES cells are dynamically heterogeneous at the population level[5]. A more appropriate approach than fitting a two-state HMM is to fit an HMM with at least three states: baseline, LOCKs and apparent LOCKs due to underlying biology. It is clear that had Filion and van Steensel jointly fitted a three-state HMM, instead of a two-state HMM, to the undifferentiated and differentiated datasets, they would have obtained results very similar to ours.

Our experimental and statistical methods extend the boundaries of our ability to define differences in nuclear organization and are imperfect, just as is van Steensel's method for defining lamin-associated domains (LADs) through *in vivo* methylation by lamin fusion proteins[6]. But if our conclusions were wrong, why would the LOCKs we defined detect regions largely overlapping with LADs? van Steensel's group described changes in LADs during ES cell differentiation in an abstract at the Cold Spring Harbor Conference on Dynamic Organization of Nuclear Function (2008). There is even older evidence, although not mapped to specific chromosomal locations, showing an increase in H3K9me2 in differentiated cells compared to undifferentiated ES cells[7]. Furthermore, Bing Ren and colleagues have confirmed our observation of large heterochromatin domains of hundreds of kilobases in size arising in differentiated ES cells from regions with bumps of a few kilobases in undifferentiated ES cells, albeit in human cells and with different heterochromatin markers (ref. 8 and B. Ren, personal communication). They also showed that partially methylated domains (PMDs), in which DNA are less methylated in fibroblasts compared to human ES cells[8], are enriched for expanded heterochromatin blocks in fibroblasts but not in ES cells. Interestingly, the LOCKs we defined in differentiated ES cells largely overlap the PMDs (**Supplementary Fig. 3**), even given that the mapping is cross-species.

Bo Wen[1,2], Hao Wu[1,3], Yoichi Shinkai[4], Rafael A Irizarry[1,3] & Andrew P Feinberg[1,2]

[1]Center for Epigenetics and [2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [3]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [4]Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto, Japan. Correspondence should be addressed to: A.P.F. (afeinberg@jhu.edu).

Note: Supplementary information is available on the Nature Genetics website.

1. Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A. & Feinberg, A.P. *Nat. Genet.* **41**, 246–250 (2009).
2. Conley, B.J. *et al. Curr. Protoc. Cell Biol.* Chapter 23 (2005).
3. Roach, M.L. & McNeish, J.D. *Methods Mol. Biol.* **185**, 1–16 (2002).
4. *Stem Cells: Scientific Progress and Future Research Directions* (US Department of Health and Human Services, Washington, DC, USA, 2001).
5. Kalmar, T. *et al. PLoS Biol.* **7**, e1000149 (2009).
6. Guelen, L. *et al. Nature* **453**, 948–951 (2008).
7. Dai, B. & Rasmussen, T.P. *Stem Cells* **25**, 2567–2574 (2007).
8. Lister, R. *et al. Nature* **462**, 315–322 (2009).

Evolutionary flux of canonical microRNAs and mirtrons in *Drosophila*

To the Editor:

Next-generation sequencing technologies generate vast catalogs of short RNA sequences from which to mine microRNAs (miRNAs), which are ~21–24-nucleotide regulatory RNAs derived from RNase III–mediated cleavages of hairpin transcripts. However, such data must be vetted to appropriately categorize miRNA precursors and interpret their evolution. A recent study annotated hundreds of miRNAs in three *Drosophila* species on the basis of singleton reads of heterogeneous length[1]. Our multimillion-read datasets indicated that most of these putative miRNAs were not produced by RNase III cleavage and that they comprised many mRNA degradation fragments. We instead identified a distinct and smaller set of new miRNAs supported by high-confidence cloning signatures, which included a high proportion of evolutionarily nascent mirtrons. Our data support a much lower rate for the emergence of lineage-specific miRNAs than was previously inferred[1], with a net flux of ~1 miRNA per million years of drosophilid evolution.

Conserved miRNA genes are differentiated from bulk hairpins in that their terminal loops diverge more quickly than their stems[2]. However, species-specific miRNAs cannot be confidently identified by using solely computational methods, as hundreds of thousands of *Drosophila*[1,3–5] and human loci[6] are plausible as miRNA hairpins. Instead, we and others have turned to next-generation sequencing to identify recently evolved miRNAs, which lack support from evolutionary signatures (for example, **Supplementary Table 1**). Such deep sequence data often reveal heterogeneous size and read patterns with respect to predicted hairpins (**Fig. 1** and **Supplementary Fig. 1**), indicating that only a subset of hairpins with reads are substrates of Dicer-driven biogenesis pathways. In particular, it is not possible to determine whether a predicted hairpin associated with a single-cloned short RNA is indeed an endogenous substrate of RNase III cleavage (**Fig. 1**).

Lu and colleagues reported ~900 putative novel miRNAs sequenced from three *Drosophila* species—*D. melanogaster* (*Dme*), *D. simulans* (*Dsi*) and *D. pseudoobscura* (*Dps*)—including ~400 annotated under ‘high-stringency’ criteria[1]. They concluded that evolutionarily transient miRNA genes are continually born and lost, with only a small proportion of miRNAs fixed across drosophilid radiation. Inspection of these annotations showed that 35 *Dme*, 47 *Dsi* and 30 *Dps* ‘novel’ miRNAs corresponded to orthologs of 50 distinct genes whose cloning and evolutionary characteristics had been previously described[4,5,7] (miRBase 10.1 and **Supplementary Tables 2–4**). Another locus comprising multiple tandem hairpins corresponded to hairpin RNA hp-CG4068, which generates endogenous small interfering RNAs (endo-siRNAs)[8]. We sought to understand the nature of the remaining hundreds of miRNA candidates, whose abundant numbers were previously used to estimate a birthrate of ~12 miRNAs per Myr of drosophilid evolution[1].

We mapped ~15 million *Dme* reads from diverse developmental stages and tissues, including ~1 million from adult heads[4,9]. Compared to their frequency among ~16,000 reads from adult *Dme* heads[1], we expected our data to contain ~60-fold more reads for genuine miRNAs and likely more, given that many are expressed in multiple stages and tissues. This was true for the 35 *Dme* miRBase 10.1 loci designated ‘novel’ by Lu and colleagues[1]. These ‘novel’ loci were represented by 1,247 reads in their data (~34 reads per locus, although 6 loci were cloned only 2–3 times and 12 were singletons) but by ~320,000 reads in our data (~8,800 reads per locus). The remaining 23 non-miRBase loci were severely under-represented in our data, with 9 cloned 1–6 times and 9 that were not recovered at all (**Supplementary Table 2**).

For non-miRBase loci cloned in our dataset, the reads mapped incoherently across the predicted hairpin and/or adjacent genomic regions (**Fig. 1** and **Supplementary Fig. 1**). They also showed broadly heterogeneous sizes, contrasting with the restricted lengths of genuine *Drosophila* miRNAs (**Fig. 2**). Although some loci were conserved, the most abundant reads mapped to a ribosomal RNA (rRNA; *Lu-mir-2018*) and two small nuclear RNAs (snoRNAs; *Lu-mir-2324* and *Lu-mir-2213*); 16 out of the 20 remaining loci derived from mRNAs (**Supplementary Table 2**). Therefore, instances of conservation were attributable to protein-coding or functional RNA status and not to evol-