fully phase-resolved sequences. These reads are then combined into longer 'virtual reads', which serve as scaffolds for the assembly.

The same effect can also be achieved using fosmid clones[5,6] (**Fig. 1**). NGS reads from pools of fosmid clones are first mapped to the reference genome, virtual reads are assembled and then variants, including structural variations, are detected. However, all of these methods rely in one way or another on the reference genome, which means that they have limited power to detect rare and complex variation.

Cao *et al.*[1] report the first *de novo* assembly of a haplotype-resolved human genome (**Fig. 1**). Here the problem is inverted and reads are first assembled *de novo* within segments, and structural variation to the reference is found by comparison, thus removing bias to the reference. To do this, they generated over half-a-million 3- to 50-kb fosmid clones and sequenced them, using a dual-barcoding approach, in pools comprising only ~33 fosmids. The resulting fosmid sequences preserve phase information much like previous methods based on indexed pools[3,4] and fosmid clones[5,6].

Next, they assembled sequences from each pool into long, haplotype-resolved "fosmid assembled haploid" (FAH) sequences. The authors then performed *de novo* assembly of previously sequenced whole genome shotgun (WGS) libraries from the same donor and used the FAHs as long reads to improve continuity between the scaffolds created from WGS *de novo* assembly. The FAHs also provide phase information for sequenced DNA because they contain multiple heterozygous genetic markers used to construct linkage maps that resolve phase orientation. The resulting assembly quality is high, with half the data in phased stretches longer than 484 kb (haplotype N50).

Cao *et al.*[1] called single-nucleotide polymorphisms (SNPs) and structural variation across each haplotype and then compared these findings to the original human reference genome. Intriguingly, 7% of the SNPs and 60% of small (<50 bp) and moderate (50–200 bp) length indels identified in this study are novel and not present in either the dbSNP or 1000 Genomes databases. Many of these variants are located in highly repetitive regions, which were not well-resolved using short read data alone. A large proportion of novel indels are validated using Complete Genomics and Illumina HiSeq2000 resequencing technologies, reflecting the power of the approach to identify structural variation that current workflows do not readily detect.

There are at least three areas in which *de novo* haplotype-resolved assemblies of human genomes will be immensely useful. First, they will facilitate the detection and characterization of complex structural variants that may have important biological consequences. Second, this technology provides a unique opportunity in clinical genetics to understand modalities of disease inheritance. Some diseases are influenced by the phase of two variants in the disease gene. For example, in the case of autosomal recessive disorders, a patient with two pathogenic variants will be predicted to either carry or express the disease depending on whether these variants occur on the same or different chromosomes, respectively.

A third field that will benefit from this technology is human population and medical genetics. Phase information is used to estimate rates of recombination, which has previously been accomplished through sperm typing[7] and family-based phasing[8]. However, these approaches are often slow and labor-intensive and can require recruitment of large families, which is nontrivial. Many tests of selection and estimators of migration also make use of haplotype information, and current methods rely almost exclusively on statistical inference based on simplistic evolutionary models. Genome phasing could provide much needed orthogonal data for validating these approaches and for understanding the recent history of rare alleles that cannot be reliably phased using statistical approaches.

The affordability and quality of *de novo* haplotype-resolved genomes will continue to improve in tandem with advances in long-read[9] and nanopore sequencing, in sample preparation for short-read sequencing and in approaches such as optical mapping. Each of these technologies has unique advantages and disadvantages, and it is likely that some combination of them will prove most successful in validating and optimizing sequence quality and assembly.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.3249).

1. Cao, H. *et al. Nat. Biotechnol.* **33**, 617–622 (2015).
2. Levy, S. *et al. PLoS Biol.* **5**, e254 (2007).
3. Amini, S. *et al. Nat. Genet.* **46**, 1343–1349 (2014).
4. Kuleshov, V. *et al. Nat. Biotechnol.* **32**, 261–266 (2014).
5. Suk, E.-K. *et al. Genome Res.* **21**, 1672–1685 (2011).
6. Kitzman, J.O. *et al. Nat. Biotechnol.* **29**, 59–63 (2011).
7. Wang, J., Fan, H.C., Behr, B. & Quake, S.R. *Cell* **150**, 402–412 (2012).
8. Roach, J.C. *et al. Science* **328**, 636–639 (2010).
9. Chaisson, M.J.P. *et al. Nature* **517**, 608–611 (2014).

Epigenome editing made easy

Gabriel E Zentner & Steven Henikoff

Fusions of Cas9 to histone-modifying enzymes enable functional interrogation of the epigenome.

The epigenome is a layer of regulatory information superimposed on the genome, comprising the positions, compositions and modifications of nucleosomes as well as modifications of DNA bases[1]. Mapping epigenomic components genome-wide in many cell types is a major ongoing effort in the epigenetics community that has already yielded millions of putative regulatory elements. But functional characterization of these elements has lagged behind their discovery owing to a lack of easy-to-use technologies for manipulating the epigenome at individual loci. Recent reports in *Nature Biotechnology*, by Hilton *et al.*[2], and in *Nature Methods*, by Kearns *et al.*[3], have now addressed this need with an epigenome editing technology based on CRISPR-Cas9 (clustered, regularly interspaced, short palindromic repeat–CRISPR-associated protein). By combining the ability of CRISPR-Cas9 to direct effectors to specific genomic sites with well-characterized, histone-modifying domains, these studies functionally characterized the chromatin states of specific genomic elements and showed that modulation of histone modifications has robust, specific transcriptional outcomes.

Previous work interrogating the roles of functional elements and their associated chromatin modifications used transcription activator-like effectors (TALEs) and zinc finger nucleases (ZFNs) to target specific sites[4–6], with promising results. For example, fusion of the lysine-specific histone demethylase LSD1 to TALEs reduced histone H3 lysine 4 dimethylation (H3K4me2) and H3K27 acetylation

Gabriel E. Zentner is in the Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, and Steven Henikoff is in the Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.
e-mail: steveh@fredhutch.org

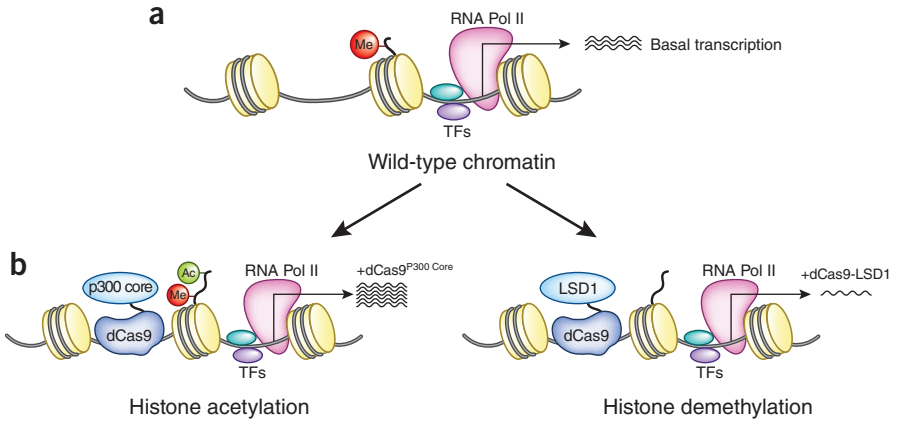

Figure 1 Cas9-based epigenome editing. (**a**) At a given promoter, H3K4me2, transcription factors (TFs) and RNA Pol II synergize to promote a basal level of gene expression. (**b**, left) Targeting of dCas9$^{P300\ Core}$ to this promoter increases H3K27ac, leading to transcriptional upregulation. (**b**, right) Targeting of dCas9-LSD1 to this promoter removes H3K4me2, leading to transcriptional downregulation.

(H3K27ac) at target sites and the expression of nearby genes[5]. However, these approaches are relatively laborious as a TALE or ZFN variant protein must be engineered for each desired target sequence, whereas Cas9 targeting is more straightforward, requiring only provision of an appropriate guide RNA.

To modulate histone acetylation at promoters and enhancers, Hilton *et al.*[2] fused a nuclease-deficient *Streptococcus pyogenes* Cas9 (dCas9) to the catalytic domain of p300 (dCas9$^{p300\ Core}$), which acetylates H3K27 (**Fig. 1**). They targeted dCas9$^{p300\ Core}$ to the promoters of *IL1RN*, *MYOD* and *OCT4* in HEK293T cells, which do not normally express these genes. To ensure robust acetylation, the authors first simultaneously expressed multiple guide RNAs (gRNAs) targeting each promoter, although later experiments showed that a single gRNA is often sufficient, further simplifying the approach. Targeting of dCas9$^{p300\ Core}$ to each promoter substantially upregulated the associated gene, even compared with a fusion of Cas9 and the VP64 acidic activation domain (dCas9-VP64).

As H3K27ac marks enhancers associated with actively transcribed genes, the authors also asked whether targeting of dCas9$^{p300\ Core}$ to previously characterized *MYOD* and *OCT4* enhancers would activate transcription of their associated genes. Indeed, this was the case for both of two tested enhancers of each gene. Similarly, localization of dCas9$^{p300\ Core}$ to the HS2 enhancer in the β-globin locus activated transcription of multiple hemoglobin genes. Expression profiling by RNA-seq revealed that dCas9$^{p300\ Core}$-mediated transcriptional activation was robust and specific, and ChIP-PCR showed high levels of H3K27ac at targeted regions. Lastly, portability of the p300 catalytic core was demonstrated with a fusion to dCas9 from *Neisseria meningitidis* (*Nm*).

To study histone methylation at enhancers, Kearns *et al.*[3] fused *Nm*-dCas9 to LSD1 (**Fig. 1**). Targeting of dCas9-LSD1 to the distal enhancers of *Oct4* and *Tbx3* in mouse embryonic stem cells strongly reduced gene expression, whereas targeting of the enzyme to the promoters had no effect on gene expression, indicating the specificity of the fusion. ChIP-PCR analysis of histone modifications at the *Tbx3* distal enhancers upon dCas9-LSD1 targeting revealed a marked decrease in H3K4me2 and H3K27ac, perhaps reflecting recruitment of the NuRD deacetylase complex by LSD1 (ref. 7), whereas histone marks at the *Tbx3* promoter were unchanged.

The results of the two studies suggest that targeted epigenome editing using CRISPR-Cas9 will be a powerful, simple method for probing the effects of histone modifications at specific loci. However, caution may be warranted when interpreting the results of such experiments, as protein-modifying enzymes often have many non-histone targets. This is exemplified by the recent finding that H4K20, thought to be essential for development and DNA replication based on studies of its modifying enzymes, did not affect either process when mutated[8].

Post-translational protein modifications are only one component of the epigenome, and it will be interesting to use CRISPR-Cas9 to functionally characterize DNA modifications. Fusion of TALEs to methylcytosine dioxygenase TET1 for targeted DNA de-methylation has been reported, so dCas9-TET1 fusions may be a fruitful direction. One interesting application could be investigation of the proposed DNA methylation–dependent enhancer memory[9]. This phenomenon could also be interrogated using fusions of dCas9 with DNA methyltransferases.

CRISPR-Cas9-based epigenome editing may also be useful in assessing the functional consequences of nucleosome occupancy, positioning and composition. Numerous genome-wide maps of these parameters have been generated, but, as with maps of histone modifications, an understanding of their effects on gene expression has remained largely correlative. Fusion of dCas9 to a variety of nucleosome-modulating factors would be a simple method to probe the role of nucleosomal parameters in transcription *in vivo*. Indeed, a recent report described the use of ZFNs to target chromatin regulatory proteins, including chromatin remodelers and histone chaperones, to a reporter locus in budding yeast[6].

Pharmaceutical inhibitors of histone-modifying enzymes have been widely used in basic research and in clinical applications. However, these treatments cause global changes in histone marks and transcription, and it is not easy to determine which of the millions of putative functional elements associated with a specific mark mediate the observed phenotypic effects. For instance, inhibition of the H3K79 methyltransferase Dot1L with the drug EPZ004777 results in a global decrease in H3K79me and widespread transcriptional alterations. EPZ004777 treatment specifically inhibits the growth of mixed lineage leukemia (MLL)-transformed leukemia cells[10], indicating that it affects MLL fusion target genes.

To find the relevant targets of Dot1L in leukemia, one could fuse dCas9 to the catalytic core of Dot1L engineered to resist EPZ004777 and express fusion protein in leukemia cells treated with EPZ004777 (that is, depleted of H3K79me) along with gRNAs targeting putative leukemogenic genes such as *HOXA9* and *MEIS1*. This approach could also be generalized to other drugs that inhibit cancer-relevant histone-modifying enzymes, such as EZH2. We expect that targeting of histone-modifying activities to specific loci with dCas9 will be a powerful tool for both understanding the nuances of the epigenome and translating this knowledge to the clinic.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1 Zentner, G.E. & Henikoff, S. *Nat. Rev. Genet.* **15**, 814–827 (2014).
2. Hilton, I.B. *et al. Nat. Biotechnol.* **33**, 510–517 (2015).
3. Kearns, N.A. *et al. Nat. Methods* **12**, 401–403 (2015).
4. Maeder, M.L. *et al. Nat. Biotechnol.* **31**, 1137–1142 (2013).
5. Mendenhall, E.M. *et al. Nat. Biotechnol.* **31**, 1133–1136 (2013).
6. Keung, A.J. *et al. Cell* **158**, 110–120 (2014).
7. Whyte, W.A. *et al. Nature* **482**, 221–225 (2012).
8. McKay, D.J. *et al. Dev. Cell* **32**, 373–386 (2015).
9. Hon, G.C. *et al. Nat. Genet.* **45**, 1198–1206 (2013).
10. Daigle, S.R. *et al. Cancer Cell* **20**, 53–65 (2011).