Genetic sources of population epigenomic variation

Aaron Taudt1, Maria Colomé-Tatché1,2 and Frank Johannes3,4

Abstract | The field of epigenomics has rapidly progressed from the study of individual reference epigenomes to surveying epigenomic variation in populations. Recent studies in a number of species, from yeast to humans, have begun to dissect the cis- and trans-regulatory genetic mechanisms that shape patterns of population epigenomic variation at the level of single epigenetic marks, as well as at the level of integrated chromatin state maps. We show that this information is paving the way towards a more complete understanding of the heritable basis underlying population epigenomic variation. We also highlight important conceptual challenges when interpreting results from these genetic studies, particularly in plants, in which epigenomic variation can be determined both by genetic and epigenetic inheritance.

DNA functionally interacts with a variety of epigenetic marks, such as cytosine methylation (also known as 5-methylcytosine (5mC)) or histone modifications (FIG. 1a). The dynamic placement of these marks along the genome is essential for coordinating gene expression programmes and for maintaining genome integrity in response to developmental or environmental cues. Technological advances in the past decade have enabled high-resolution measurements of various epigenetic marks at a genome-wide scale5–7 (FIG. 1b). The computational integration of these measurements has led to the construction of so-called chromatin state maps (FIG. 1c), which provide an operational definition for the term ‘epigenome’. These integrated maps are believed to give a good description of the functional state of the genome in a given cell type and at a specific time-point. Large initiatives are underway to collect reference epigenomes for different tissues, developmental stages, disease states and environmental treatments8–10. This information has already been instrumental in elucidating key chromatin changes during cellular differentiation, disease pathology and for functionally annotating causal variants from human genome-wide association mapping studies8–10.

Reference epigenomes are usually derived from cells of a single individual or from a pool of several individuals and therefore do not capture inter-individual epigenomic variation at the population level. Genetic polymorphisms or differential environmental exposure can alter chromatin states and lead to transient or permanent changes in gene expression. Chromatin states therefore represent important molecular phenotypes that mediate how different genotypes are translated into observable traits or how environmental signals are translated into genomic function. There is substantial interest in the biomedical, agricultural and evolutionary communities to try to understand the factors that cause population epigenomic variation. Various epidemiological studies have begun to address this problem by searching for specific environmental causes8–10. Complimentary to these approaches, a number of recent genetic studies have tried to quantify the heritable basis underlying population epigenomic variation and to delineate its cis- and trans-regulatory architecture (FIG. 1d; TABLE 1).

These recent genetic studies vary widely in scope: they consider different species, sample sizes, methodological approaches, measurement technologies and units of analysis. Here, we provide a critical review of these emerging efforts. Our Review focuses on the proportion of the epigenome that is found to be under genetic control, the relative contributions of cis- and trans-acting factors, their average effect sizes and mechanisms of action. We also highlight important conceptual and technical issues in the construction of chromatin state maps and in the interpretation of genetic associations detected in these studies, particularly in plants, in which epigenomic variation can be determined both by genetic and epigenetic inheritance. Scaling up current studies to include more epigenetic marks, cell types and individuals promises to provide deeper insights into the heritable basis underlying population epigenomic variation and will clarify its implications for biomedical, agricultural and evolutionary research.

Chromatin state maps define epigenomes
Genomic DNA is tightly packed in cells, and the basic unit of DNA packaging is called the nucleosome.
Approximately 150 bp of DNA wrap around a histone octamer, which consists of two copies of each of the core histones (H2A, H2B, H3 and H4). In addition to direct modifications of DNA in the form of 5mC, core histones can be subjected to a variety of chemical modifications of their amino acid residue tails\textsuperscript{10} (FIG. 1a). Genome-wide maps of 5mC and various histone modifications can be readily obtained with array or next-generation sequencing (NGS) technologies coupled with bisulfite conversion or immunoprecipitation assays\textsuperscript{1,2} (FIG. 1b).

To date, more than 100 histone modifications have been described\textsuperscript{11,12}. This large number has led to the idea of an epigenetic code\textsuperscript{13,14} — a layer of information that is encoded by recurring patterns of epigenetic marks. This

![Figure 1](image-url)

**Figure 1 | Main steps in population epigenomic analysis.** a | DNA is tightly packaged in cells and is functionally modified by a variety of epigenetic marks, such as cytosine methylation (5mC) or post-translational changes in histone proteins. The co-occurrence of specific epigenetic marks in a genomic region defines its functional state. Of note, histones in closed chromatin also contain repressive marks (not shown). b | The genome-wide distribution of different epigenetic marks can be measured using next-generation sequencing (NGS) technologies. Shown are the read-tracks from NGS measurements of N different epigenetic marks along the genome. c | The computational challenge is to infer distinct chromatin states for each genomic position. These chromatin states are defined by the joint presence and absence patterns of the different epigenetic marks. With N marks there can be $2^N$ possible combinatorial states. The colour code on the bottom denotes each unique state. This analysis leads to the construction of chromatin state maps. d | Shown are the chromatin state maps of M diploid individuals. Individuals differ in their chromatin states in three genomic regions. These differential chromatin states (DCSs) can originate from DNA sequence polymorphisms, environmental factors or from stochastic changes. DCS2 is caused by a single-nucleotide polymorphism (SNP2), DCS3 is caused by exposure to environmental factor E4 and DCS1 is the result of stochastic processes in the mitotic maintenance of the chromatin state at that locus. The statistical challenge is to try to identify these causal factors from millions of measured SNPs and a large number of environmental factors.
<table>
<thead>
<tr>
<th>Organism</th>
<th>Population</th>
<th>Tissue</th>
<th>N ind.</th>
<th>N marks</th>
<th>Method</th>
<th>FDR control</th>
<th>cis (kb)</th>
<th>% assoc.</th>
<th>% cis</th>
<th>% trans</th>
<th>% var. cis</th>
<th>% var. trans</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Integrative chromatin states</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>70</td>
<td>1††</td>
<td>GWA-cis</td>
<td>0.1</td>
<td>20</td>
<td>0.64</td>
<td>100</td>
<td>NT</td>
<td>~30–40</td>
<td>NT</td>
<td>35</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>14</td>
<td>5</td>
<td>AT</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>100</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>36</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>64</td>
<td>5</td>
<td>Other method</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>100</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>7</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>10</td>
<td>4</td>
<td>GWA-cis</td>
<td>0.2</td>
<td>1</td>
<td>NR</td>
<td>100</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>37</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>47</td>
<td>3</td>
<td>GWA-cis</td>
<td>0.1</td>
<td>250</td>
<td>8.1†</td>
<td>100</td>
<td>NT</td>
<td>~40</td>
<td>NT</td>
<td>38</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>75</td>
<td>3</td>
<td>GWA-cis</td>
<td>0.1</td>
<td>2§</td>
<td>10–15†§</td>
<td>100</td>
<td>20950‡</td>
<td>NR</td>
<td>NR</td>
<td>39</td>
</tr>
<tr>
<td>Rat</td>
<td>RIL</td>
<td>Heart and liver</td>
<td>30</td>
<td>2</td>
<td>LM</td>
<td>0.05</td>
<td>10§</td>
<td>NR</td>
<td>69.24, 30.76, 24.72††</td>
<td>NR</td>
<td>NR</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Yeast</td>
<td>RIL</td>
<td>NA</td>
<td>96</td>
<td>1†</td>
<td>LM</td>
<td>0.01</td>
<td>100</td>
<td>NR</td>
<td>7.9</td>
<td>92.1</td>
<td>NR</td>
<td>33</td>
<td>47</td>
</tr>
<tr>
<td>Yeast</td>
<td>RIL</td>
<td>NA</td>
<td>94</td>
<td>1†</td>
<td>LM</td>
<td>0.01</td>
<td>100</td>
<td>NR</td>
<td>15.7</td>
<td>84.3</td>
<td>NR</td>
<td>NR</td>
<td>48</td>
</tr>
<tr>
<td><strong>DNA methylation in humans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>CRBLM</td>
<td>150</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.1</td>
<td>10§</td>
<td>4–5.1</td>
<td>30–40</td>
<td>60–70</td>
<td>18–88††</td>
<td>18–88††</td>
<td>65</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>153</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>10§</td>
<td>8.71§</td>
<td>98</td>
<td>2</td>
<td>17–73</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>77</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.1</td>
<td>50</td>
<td>0.17</td>
<td>73</td>
<td>27</td>
<td>22–63</td>
<td>NR</td>
<td>67</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>180§</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.3–0.4</td>
<td>100</td>
<td>0.12–0.38</td>
<td>100</td>
<td>NT</td>
<td>36–92†§</td>
<td>NT</td>
<td>68</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>WB</td>
<td>201</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>100</td>
<td>11.93§</td>
<td>100</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>69</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>133§</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>100</td>
<td>5.1–5.8</td>
<td>100</td>
<td>NT</td>
<td>30†§†</td>
<td>NT</td>
<td>70</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>FIB, T cells and LC</td>
<td>66–111</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.1</td>
<td>5</td>
<td>3.4–7.8</td>
<td>100</td>
<td>NT</td>
<td>10–90</td>
<td>NT</td>
<td>71</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>HPI</td>
<td>89</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>500</td>
<td>2.57</td>
<td>96.8</td>
<td>3.2</td>
<td>NR</td>
<td>NR</td>
<td>72</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>FIB</td>
<td>62</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>250</td>
<td>2§</td>
<td>100</td>
<td>NR</td>
<td>NT</td>
<td>NT</td>
<td>73</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>133§</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>10§</td>
<td>1.8–2.6</td>
<td>100</td>
<td>NT</td>
<td>23–97†§</td>
<td>NT</td>
<td>74</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>LCL</td>
<td>34§</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.05</td>
<td>200</td>
<td>NR</td>
<td>100</td>
<td>NT</td>
<td>26§</td>
<td>NT</td>
<td>64</td>
</tr>
<tr>
<td>Human</td>
<td>Natural</td>
<td>WB</td>
<td>697</td>
<td>1</td>
<td>GWA-cis</td>
<td>0.01</td>
<td>10§</td>
<td>15</td>
<td>99.4</td>
<td>0.6</td>
<td>NR</td>
<td>NR</td>
<td>75</td>
</tr>
<tr>
<td><strong>DNA methylation in plants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>Natural and RIL</td>
<td>Leaf</td>
<td>51</td>
<td>1</td>
<td>GWA-cis and LM</td>
<td>0.05</td>
<td>2</td>
<td>51</td>
<td>100</td>
<td>NT</td>
<td>NR</td>
<td>NT</td>
<td>83</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>Natural and RIL</td>
<td>Leaf and MSIT</td>
<td>152</td>
<td>1</td>
<td>GWA</td>
<td>0.01</td>
<td>NR</td>
<td>16–35</td>
<td>26</td>
<td>74</td>
<td>NR</td>
<td>NR</td>
<td>81</td>
</tr>
<tr>
<td>Soybean</td>
<td>RIL</td>
<td>Leaf</td>
<td>83</td>
<td>1</td>
<td>LM</td>
<td>NR</td>
<td>CHR</td>
<td>91</td>
<td>97.5</td>
<td>2.5</td>
<td>51§†§</td>
<td>51§†§</td>
<td>84</td>
</tr>
<tr>
<td>A. thaliana</td>
<td>Natural</td>
<td>WR</td>
<td>155</td>
<td>1</td>
<td>GWA</td>
<td>0.05***</td>
<td>100</td>
<td>21†††</td>
<td>18†††</td>
<td>31†††</td>
<td>41†††</td>
<td>60†††, 55†††</td>
<td>-10*†††</td>
</tr>
</tbody>
</table>

A. thaliana, Arabidopsis thaliana; Assoc., association; AT, allele transmission mapping; CHR, chromosome; CRBLM, cerebellum; FCTX, frontal cortex; FDR, false discovery rate; FIB, fibroblasts; GWA, genome-wide association mapping; GWA-cis, genome-wide association mapping that tests only for associations in cis; HPI, human pancreatic islets; ind., individuals; LC, lymphoblastoid cells; LCL, lymphoblastoid cell lines; LM, linkage mapping; MSIT, mixed-stage inflorescence tissue; NA, not available; NR, not reported; NT, not tested; RIL, recombinant inbred lines; TCTX, temporal cortex; WB, whole blood; WR, whole rosettes; var., variation. *Studies were selected if they included more than 10 individuals, used genome-wide methods for measuring epigenomic variation and applied mapping approaches to identify cis- and/or trans-acting genetic variants; a larger table with additional details on the types of marks studied is available as Supplementary information S3 (table). †Measurement of open chromatin only. ‡Data contains individuals from diverse populations. §Number of distal quantitative trait loci (QTL) found when conditioning on detected modules that show association. ††Trans is defined as >50 kb and <2 Mb. †‡Percent of associations in cis. †††Number of distal quantitative trait loci (QTL) found when conditioning on detected QTL; note that this number is not a percentage. ¶Heart and liver, respectively. †‡‡Percent of variance explained by cis and trans loci combined. †‡‡†Reported numbers refer to a conditional analysis in which variable probes were pre-selected. †‡‡‡Numbers are based on within-population analysis. †††Average estimate. ††††FDR was only applied to CHH methylation analysis; reported number for gene-body methylation is based on results without FDR control. †††††Refers to the analysis of gene body methylation. ††††††Refers to the analysis of CHH methylation.
REVIEWS

Box 1 | Definitions of chromatin states

Since the proposition of the existence of a ‘histone code’ in 2000 (REFS 13,14), considerable effort has been spent to decipher this code, and many computational approaches have been developed to integrate single marks into chromatin state maps. Different conceptual ideas of a chromatin state underlie the different approaches. The original notion of a histone code11,14 is based on a molecular view that assumes that histone modifications (or epigenetic marks in general) are either present or absent at any given position in the genome in a binary manner, so that their combined presence and absence patterns define distinct combinatorial chromatin states15,24,28,13,137.

A second view takes into account the continuous nature of the ChIP–seq (chromatin immunoprecipitation followed by sequencing) signal and defines chromatin signatures on the basis of the signal shape rather than by the binary presence or absence of every mark18,138. A third view defines probabilistic chromatin states6,19,20,23,26 (also called fuzzy chromatin states20), which have probabilities associated with finding each mark in a given state, meaning that one state can be a superposition of multiple combinatorial patterns with different probabilities. A fundamental problem of chromatin-state-calling algorithms is to infer the ‘true’ number of states. Although it is reasonable to assume that the number of states increases with the number of epigenetic marks, our review of the literature shows that there is no clear trend (see the figure, part a; Supplementary information S1 (table)). There are several reasons for this: first, different experimental techniques and analytical approaches investigate the epigenome at different resolutions, with higher resolution potentially leading to more chromatin states; second, the number of chromatin states is a function of the investigated marks (a set of uncorrelated marks has more states than a set of correlated or redundant marks); third, the majority of computational methods treat the number of chromatin states as an input rather than an output of the analysis, so that chromatin states reflect previous knowledge of chromatin. Another interesting question is the percentage of the genome that is covered with epigenetic marks or, conversely, is devoid of any marks. Our review of the literature shows that the percentage of empty epigenome decreases when more marks are measured (see the figure, part b; Supplementary information S1 (table)). Indeed, one experiment involving 53 marks by Filion et al.23 showed that essentially no part of the genome is permanently without epigenetic modifications.

\[
\begin{align*}
\text{Algorithm} & \quad \text{ChromaSig} \\
& \quad \text{ChromHMM} \\
& \quad \text{ChromstaR} \\
& \quad \text{Clustering} \\
& \quad \text{Nucleosome alphabet} \\
& \quad \text{Principal component analysis} \\
& \quad \text{Post-hoc combination} \\
& \quad \text{Segway}
\end{align*}
\]

Population genetics of epigenomes

Our knowledge of the extent to which tissue-specific combinatorial chromatin states are variable at the population level and the extent to which they are influenced by genetic variation is mainly limited to a few small-scale studies in humans25,39 (TABLE 1). These studies profiled several common histone modifications in lymphoblastoid cell lines of individuals whose genomes were also sequenced as part of the 1000 Genomes Project. The modifications include mostly active marks (monomethylated histone H3 lysine 4 (H3K4me1), trimethylated H3K4 (H3K4me3), acetylated H3K27 (H3K27ac), H4K20me1, and H3K36me3), but also one repressive mark (H3K27me3), and were chosen on the basis of their important role in determining tissue- and development-specific gene expression programmes40. The genomic distribution of these marks is heavily biased toward genes. On average, ~63% of the modifications map within or in close proximity to exons (<5kb) and together cover 25% of the total human genome (numbers were generated based on data from REF. 7).
**Genetic effects on chromatin states in cis.** Kasowski et al. used 5 of these histone modifications to construct integrated chromatin state maps in 19 individuals. Using probabilistic chromatin states as units of analysis, they showed that genomic regions corresponding to bivalent enhancers (defined by the authors as having H3K4me1 and H3K27me3), weak enhancers (defined as H3K4me1) and weak enhancers in transcribed regions (defined as H3K4me1 and H3K36me3) were most variable between individuals. Inter-individual differences were mainly in the form of state transitions from an enhancer state to a repressive state (for example, Polycym-associaited state H3K27me3 or empty state; see BOX 1), or vice versa. State transitions of this type reflect what is typically seen during cellular differentiation, suggesting that intra- and inter-individual epigenetic variation converge on the same genomic regions. Moreover, changes in H3K4me1 density have been found to be the best predictor of changes in long-range chromatin interactions, indicating that enhancer variation may also translate into variation in higher-order chromatin structure.

Focusing on a subset of these histone marks (TABLE 1), a number of larger studies (involving up to 75 individuals) used association mapping approaches to identify cis-regulatory single-nucleotide polymorphisms (SNPs) — that is histone quantitative trait loci (hQTL) — within several kilobases of these modifications. Rather than working with combinatorial states, they analysed each histone modification separately. Only a small proportion of the tested sites seem to be associated with local SNPs (1.1–15%; TABLE 1), whereas causative variants often affect multiple histone modifications at once in a manner that is consistent with their combinatorial occurrence in the genome (FIG. 2a, b). Perhaps not unexpectedly, the effects of these hQTL are not limited to histone modifications but extend to other correlated molecular phenotypes, such as local nucleosome positioning and chromatin accessibility. Such coordinated, multi-layered chromatin changes can span several thousand base pairs and provide a mechanism by which hQTL act as expression quantitative trait loci (eQTL) to proximal genes without actually being located in coding regions.

Waszak et al. and Grubert et al. recently demonstrated that similar coordinated changes also explain more distal eQTL that influence their target genes from several megabases away (FIG. 2c). They superimposed hQTL and eQTL results on a reference Hi-C map and found that distal associations coincide with long-range chromatin contacts between distal enhancers and proximal promoters residing within the same topological domain. Chromatin states at these contact points are highly correlated probably owing to their physical proximity. As a result of these state correlations, hQTL at distal enhancers tend to also associate with chromatin states at proximal promoters as well as with expression levels of genes corresponding to these promoters (FIG. 2).

The mechanisms by which single SNPs can achieve such complex cis-regulatory control remain largely elusive. The most promising model posits that SNP alleles disrupt transcription factor binding sites (TFBSs) in enhancers and/or promoters and thereby induce allele-specific changes in the chromatin environment. This causal model is supported by direct experimental evidence showing that the de novo insertion of transcription factor binding motifs is sufficient to establish active chromatin states, at least for some classes of transcription factors. However, TFBS-disrupting SNPs seem to account for only a fraction of the detected hQTL; the remainder require further experimental molecular analysis.

**Genetic effects on chromatin states in trans.** With sample sizes of up to only 75 individuals, these pioneering studies of integrated chromatin state maps were clearly underpowered to extend association analysis beyond cis-regulatory regions. Mutations in trans-acting transcription factors or in chromatin control genes, such as histone methyltransferases, may be an important source of population epigenomic variation. Numerous clinical studies report genome-wide alterations in DNA methylation and histone modification patterns in cancer cells; these alterations frequently co-occur with somatic mutations in key chromatin control genes. Mice with mutations in histone methyltransferase display similar widespread epigenomic dysregulation and highly deleterious phenotypes. However, weak alleles of these trans-acting factors can be viable and potentially segregate in populations.

Evidence for this comes from population epigenomic analyses in rats and yeast. Rintisch et al. profiled H3K4me3 and H3K27me3 in heart and liver tissue from recombinant inbred lines (RILs) derived from two different inbred rat strains. Using linkage analysis, they showed that up to 30% of all detected associations were in trans (>10 Mb). One notable example was a trans-acting hQTL that affected histone modification states at 833 target locations throughout the genome. This hQTL hotspot contained WD repeat domain 5 (WDR5), a protein-coding gene required for global and gene-specific K4me3 (REF. 40), which could potentially account for these pleiotropic effects. Similarly pronounced trans associations have been observed in QTL studies of open chromatin regions (OCRs), a proxy for active histone marks, in yeast RILs. Trans-acting loci (>100 kb) accounted for the majority (~88%) of all SNP-OCR associations, explained on average 30% of the variation in OCR patterns and were enriched for chromatin remodelers and transcription regulators. These results suggest that a systematic analysis of the trans effects on integrated chromatin states in humans may uncover additional genetic determinants.

The fact that cis- and trans-acting SNPs influence chromatin state variation implies that this variation is, at least partly, heritable at the population level (BOX 2). Genome-wide surveys of several human parent–offspring trios show that the patterns of open chromatin and histone modifications in lymphoblastoid cell lines are more similar among related individuals than unrelated individuals. Beyond these descriptive observations there are currently no estimates of the proportion of chromatin state variation that can be attributed to heritable and non-heritable sources.
A more quantitative understanding of the heritable basis of population epigenomic variation comes from more focused studies of DNA methylation in humans and plants. These studies employ much larger sample sizes, which enable robust statistical inferences.

**Population genetics of 5mC in humans**

Cytosine methylation is a widely conserved epigenetic mark with major roles in the regulation of gene expression and the silencing of transposable elements and repeat sequences. In humans, the majority

---

**Figure 2** | **Single-nucleotide polymorphisms affecting chromatin states produce signatures of molecular pleiotropy.**

**a** | Shown are the next-generation sequencing read-tracks of histone modifications: monomethylated histone H3 lysine 4 (H3K4me1), acetylated H3K27 (H3K27ac), trimethylated H3K4 (H3K4me3) and H3K27me3 for three diploid individuals with single-nucleotide polymorphism (SNP) genotypes (AA, AB and BB). Specific combinations of these epigenetic marks in a given genomic region define distinct chromatin states. The combination of H3K4me1 and H3K27ac defines 'active enhancers', H3K4me1 alone defines 'weak enhancers', H3K4me1, H3K4me3 and H3K27ac together define 'active promoters', and the absence of any measured mark defines an 'empty state'. The SNP alleles (A and B) simultaneously affect the presence and absence patterns of multiple marks. These pleiotropic effects manifest as haplotype-specific chromatin states (bottom panel).

**b** | The SNP induces chromatin state changes both in cis (within ~50 kb) and at a distal location (~1 Mb). The AA genotype is homozygous for 'active enhancer' states in cis and homozygous for the 'active promoter' state at the distal locus. For the heterozygous AB genotype the 'active enhancer' state is associated with the A allele in cis and the 'active promoter' state at the distal locus; the B allele is associated with the 'empty state' both in cis and at the distal locus. Finally, the BB genotype is homozygous for 'empty state' in cis as well as at the distal locus. c | The cis and distal regions interact through chromatin looping in a haplotype-specific manner. Interactions between 'active enhancers' (in cis) and distal 'active promoters' lead to haplotype-specific gene expression. In the absence of these interactions, gene expression is blocked. It is through these interactions that the SNP has molecular pleiotropic effects on chromatin states (both in cis and at the distal locus) and gene expression levels.
CpG islands
Genomic regions with a high percentage of CpGs and a large observed-to-expected CG ratio.

(70–80%) of all CpG dinucleotides are methylated across tissues, with unmethylated CpGs being mainly confined to CpG islands in promoter-proximal regions. However, the methylation status of ~20% of all CpGs is dynamically modified during cellular differentiation, mainly at distal enhancers, and contributes to tissue-specific gene expression programmes. These changes are coordinated at the level of chromatin states and involve substantial crosstalk with various histone modifications, such as H3K9me and H3K4me.

Genetic effects on 5mC in cis and trans: arrays.
Population-level studies of DNA methylation in humans have heavily relied on the Illumina Infinium HumanMethylation450 BeadChip (450k) array or its predecessor the 27k array. These platforms provide cost-effective measurements of CpG methylation in large samples. The 450k array surveys 1.5% of the 28 million CpGs in the human genome. Over 85% of measured sites fall into genes and promoter-proximal regions and cover nearly all CpG islands. In one of the largest genetically informative studies to date, McRae et al. used the 450k array to profile DNA methylation in blood samples of 614 individuals from 117 families consisting of twin pairs, their parents and siblings. Treating methylation levels at individual CpG sites as quantitative traits, they estimated narrow-sense heritability values ranging from 0 to 0.95 across the genome (mean ~0.20), which is roughly consistent with smaller studies using different cell types. Although this estimate may seem low, it is higher than the average heritability estimates obtained for gene expression levels, indicating that inter-individual differences in CpG methylation are under stronger genetic control.

Box 2 | Sources of population epigenomic variation
Epigenomic variation at a locus can be treated as a quantitative trait. Heritability estimates can be obtained using classical variance components analysis using pedigree data (for example, parent–offspring, twins, and so on). In the absence of epigenetic inheritance, a non-zero heritability estimate ($h^2 > 0$) implies that epigenomic variation at the locus is under genetic control by cis- or trans-acting sequence variants. Those variants should be detectable using association or linkage mapping methods, barring complicated genetic architectures. When epigenomic variation is not heritable ($h^2 = 0$), variation could be the result of differential exposures to past or current environmental factors. Systematic identification of such environmental factors should be possible and is one goal of epigenome-wide association studies (EWAS). In the absence of causative environmental factors, epigenomic variation may be the outcome of stochastic somatic epimutations that lead to intra-individual tissue heterogeneity and inter-individual ‘epigenetic drift’. Detection of such somatic epimutations will require advances in single-cell epigenomic sequencing technologies.

In plant systems, epigenetic inheritance is well documented, which complicates the interpretation of detected genetic effects (see the figure, green boxes). For instance, detected cis associations do not necessarily imply genetic regulation but may simply be due to linkage disequilibrium (LD) between segregating epigenetic variants (epialleles) at the locus and sequence alleles of proximal single-nucleotide polymorphisms (SNPs; see BOX 3 for more details). Conversely, a lack of cis association in combination with non-zero heritability estimates could suggest that epialleles are heritable but have become disassociated from their underlying DNA sequence haplotypes as a result of high epimutation rates. Unlike in mammalian systems, germine epimutations are frequent and stable enough in plants to provide a reservoir for heritable epigenomic variation (see the figure, green dashed lines). The often stated conclusion that epigenomic variation is under genetic control whenever cis-SNP associations are detected, or non-zero heritability estimates are found, is strictly only valid if epigenetic inheritance can be assumed absent. This assumption should always be checked against emerging experimental data.
Several association mapping approaches have tried to identify specific cis- and trans-acting SNPs that account for the heritability in CpG methylation (that is, methylation QTL (meQTL)) \(^{42,43-75}\). Despite major differences in cell types, data processing and analytical techniques, a remarkably consistent picture is emerging. One consistent observation is that only a modest proportion of all surveyed CpGs can be associated with meQTL at the genome-wide scale (0.12–15%). Of these meQTL, over 90% are strictly local and affect their target sites from within several kilobases. The effect sizes of these meQTL are, however, considerable — they account for 10–97% of the variation in CpG methylation. Interestingly, genome-wide analysis shows that cis-meQTL are depleted in CpG islands compared to other regions \(^{42,66,71,73,74,75}\). This mirrors the distribution of site-specific heritability estimates, which are relatively low in CpG islands and higher with increased distance from CpG islands \(^{66}\). The main reason for this seems to be that CpG islands are constitutively hypomethylated across individuals, so there is little variation to be explained by genetic or environmental factors. This lack of variation is not only visible within populations but also between human populations \(^{64,68,70,71}\), suggesting that there are strong evolutionary constraints in the maintenance of CpG island-specific methylation levels.

By contrast, the most dynamic and variable methylated regions tend to lie outside of CpG island \(^{66,73}\), often in regions that are poorly surveyed by the 450k array \(^{66,67}\), such as genic, active and weak enhancers annotated by the Epigenome Roadmap Project (Supplementary information S2 (figure)). This observation raises the question whether current array technologies provide a representative picture of population-level methylene variation and its underlying genetic architecture. A very recent array-based platform, the Infinium MethylationEPIC BeadChip array, promises to mitigate these issues by surveying 850,000 CpG sites, including some annotated enhancers \(^{78}\). Ultimately, population-level sequencing approaches are required to gain deeper insights into the utility of these array platforms; several such approaches are emerging \(^{75,76}\).

**Genetic effects on 5mC in cis and trans: NGS.** Recently, McClay et al. \(^{75}\) used methyl-CpG-binding domain (MBD) protein-enriched genome sequencing (MBD-seq) to determine the blood methylomes of 697 individuals. They interrogated ~3.2 million CpG sites throughout the genome, 15% of which could be associated with meQTL. This is about a two- to threefold increase compared to results from array-based studies, suggesting that genetic effects are much more prevalent than previously appreciated, and that current heritability estimates for CpG methylation are strongly biased downwards.

Similar to what has been reported in array-based studies, nearly all of the detected meQTL map in cis (Table 1). An estimated 75% of these cis-meQTL seem to involve simple mutations in the CpG dinucleotides themselves (that is, CpG-SNPs), thus compromising their methylation potential. This observation is consistent with an earlier chromosome-wide survey of allele-specific methylation across 16 human cell lines using base-resolution measurements \(^{79}\). The putative CpG-SNPs identified by McClay et al. \(^{75}\) are mainly located in Epigenome Roadmap Consortium annotated heterochromatin (enriched in H3K9me3) and ‘quiescent’ regions (devoid of any measured mark) \(^{79}\), and they are probably not functional. However, a considerable subset do correspond to active chromatin states, such as weak enhancers (H3K4me1), active enhancers (H3K4me1 and H3K27ac), active TSSs (H3K4me3 and H3K27ac), and show significant enrichment for genome-wide association study (GWAS) variants within 200 bp of meQTL. A similar enrichment for active chromatin was seen in the remaining 25% of cis-meQTL that did not involve obvious CpG-SNPs.

It is likely that meQTL that are associated with active chromatin tag regulatory events and correlate with local variation in other epigenetic marks. The only study to date that could assess this directly is that by Banovich et al. \(^{82}\), who integrated 450k methylation data with histone modification measurements of the same individuals. Indeed they observe pleiotropic effects for meQTL on several histone modifications as well as on proximal gene expression levels. In particular, 25% of the detected eQTL (146 eQTL; false discovery rate (FDR) = 10% \(^{83}\)) were also called as meQTL and, in half of them, gene expression and methylation levels were positively correlated. For the hQTL, 40% and 48% of those associated with H3K27ac and H3K4me3, respectively, were also classified as meQTL (FDR = 10% \(^{83}\)). This shows that these meQTL represent one facet of highly orchestrated genetic effects on local chromatin organization.

The causal mechanisms underlying cis-meQTL in regulatory regions are difficult to establish from observational data, but are probably driven by differential transcription factor binding, similar to that described above for genetic effects on integrated chromatin state. McClay et al. \(^{75}\) find a highly significant overlap between the meQTL and the binding sites of the majority of the 100 transcription factors that were profiled as part of the Epigenome Roadmap Consortium \(^{8}\). However, it remains unclear whether the data show that these transcription factors also exhibit haplotype-specific binding. This could only be properly assessed if transcription factor binding measurements were available for the same individuals that were used in the meQTL studies. Using computational predictions, Banovich et al. \(^{82}\) estimate that TFBS-disrupting SNPs account for at most 15% of detected meQTL. However, differential transcription factor binding does not necessarily require mutations in TFBSs themselves. Recently, Domcke et al. \(^{80}\) described a class of methylation-sensitive transcription factors that bind only unmethylated motifs. In this case, differential transcription factor binding may be a by-product of polymorphisms in the recognition sequences of methyltransferases or other binding proteins that disrupt maintenance methylation across larger genomic regions, thus affecting the methylation status of transcription factor binding motifs. Interestingly, NRF1 (encoded by the nuclear respiratory factor 1 gene) is one...
such methylation-sensitive transcription factor, and its target sites seem to be enriched proximal to the meQTL reported by McClay et al.26 It remains to be seen whether such events can provide the missing mechanisms underlying many of the detected cis-meQTL.

Delineating the functional basis of cis-meQTL is arguably even more challenging in plant populations, as differential methylation states can be inherited across generations independently of cis- and trans-acting sequence variants (BOX 2). As we will see, the presence of epigenetic inheritance has implications for pinpointing specific regulatory mechanisms underlying cis associations and raises broader questions regarding the heritable basis of population epigenomic variation in plants.

**Population genetics of 5mC in plants**

Population genetic studies of DNA methylation in plants have been conducted in Arabidopsis thaliana81,82, maize83 and soybean84. Although plant species differ widely in their total methylation content83, most likely owing to differences in genome size and organization85,86, patterns of intra-specific methylation variation seem to be broadly conserved81,84,87–89. Gene-rich euchromatic regions tend to be the most variable, whereas variation in transposable element (TE)-rich heterochromatic regions is largely suppressed81,83,81,90,93. The lack of variation in heterochromatic regions is consistent with robust silencing of TE sequences by small-RNA-directed mechanisms82. Unlike mammalian genomes in which methylation differences at single CpG sites can have functional consequences89, no such effects have been documented in plants. Population variation in DNA methylation is therefore usually studied at the level of differentially methylated regions (DMRs) or average methylation levels of various annotation units, as these seem to be more functional. Many mammalian studies also use the concept of DMRs; however, in the context of the genetic studies in mammals reviewed here, the units of analysis are typically individual CpGs rather than DMRs.

**Genetic effects on 5mC in cis**. In one of the first population epigenomic studies in plants, Schmitz et al.85 identified DMRs from whole-genome bisulfite sequencing data of 155 A. thaliana worldwide natural accessions (strains) and integrated this data with the full DNA sequences of the same lines. Clustering accessions based on DMRs grouped them according to genetic distance, an observation also made in maize85. One interpretation of this result is that DMRs are under strong genetic control. Using genome-wide mapping analysis, 35% of the DMRs could be associated with meQTL, with 26% of all associations mapping in cis (within 100kb). Slightly more prevalent cis effects (31–45% of all associations) were reported by Dubin et al.86, who analysed a similarly sized sample of A. thaliana natural accessions from the north and south of Sweden, although the authors used very different definitions of DMRs. In contrast to A. thaliana, cis associations seem to be far more frequent in natural populations of maize83 and RILs of soybean84. However, these latter studies either did not explicitly test for trans associations or used very liberal criteria for cis associations, which included the entire chromosome, making comparisons between genetic architectures difficult.

An emerging view suggests that many of the detected cis associations in plant populations are due to SNP alleles tagging nearby structural variants, such as TE insertions or repeats, that spread DNA methylation into flanking regions or facilitate siRNA-mediated silencing of downstream homologous sequences83,85. These structural variants not only affect DNA methylation but also establish allele-specific repressive chromatin states. Spreading of DNA methylation from TE insertions into flanking genes has been identified as a common mechanism by which TEs can drive both adaptive and non-adaptive gene expression changes86,87. Interestingly, the spreading of DNA methylation from structural variant alleles seems to be partly stochastic and thus varies between individuals both in extent and stability84,85. This stochasticity could account for the fact that detected cis-meQTL explain, on average, only ~40% of the variation in DMRs (FIG 2).

Estimates in A. thaliana and maize suggest that about 20% and 50% of all cis-meQTL are attributable to flanking structural variants, respectively83,84. The regulatory mechanisms underlying the remaining cis associations remain elusive. One possibility is that a subset of cis effects is due to TFBS-disrupting SNPs, similar to what is observed in mammalian systems. Surprisingly, there has been no systematic effort, to date, to explore this possibility in plant population epigenomic studies. In A. thaliana, this shortcoming may be due to the relatively high gene density per linkage disequilibrium (LD) block, which makes it difficult to pinpoint specific causal transcription factor binding motifs either by computational predictions or chromatin immunoprecipitation followed by sequencing (ChiP–seq). Another possibility for the lack of regulatory explanations is that cis associations in plant populations may not involve any type of genetic regulation at all, but are simply a by-product of LD between SNP alleles and segregating, meiotically stable, methylation variants (epialleles). From association or linkage mapping results alone, it is impossible to distinguish such cases of (passive) LD from active regulation, unless epigenetic inheritance can be assumed absent or epialleles are known to be highly unstable (BOXES 2,3).

**Genetic effects on 5mC in trans**. Forward and reverse genetic screens in A. thaliana and maize have identified many strong trans-acting mutations in chromatin control genes that affect DNA methylation levels genome-wide88,89. These mutants have been instrumental for delineating the molecular pathways that govern de novo and maintenance methylation in different sequence contexts. Although many of these mutants show relatively low fitness in the laboratory, mutant alleles for some of these genes seem to segregate in natural populations, and have been recovered as trans-acting meQTL82,91,100. An instructive example comes from the mapping
Box 3 Population epigenomic consequences of epimutations

Four diploid individuals sampled from the population at two different time points are shown (see the figure, part a). A meiotically stable differentially methylated region (DMR) regulates the expression of a downstream gene. The DMR is in linkage disequilibrium (LD) with a single-nucleotide polymorphism (SNP) at time $t_\alpha$, that is, SNP allele A is on the same haplotype as epiallele M, and B is on the same haplotype as U. In this case, the SNP will be detected as a methylation quantitative trait loci (meQTL), an expression QTL (eQTL) and possibly also as a QTL for higher-order complex traits, denoted here as a phenotype QTL (phQTL), without the SNP having any regulatory role in determining methylation, expression or phenotypes (see the figure, left panels of parts b and c). That is, all detected associations are simply a by-product of LD and incorrectly reflect the underlying biological reality (see the figure, part c). Since epialleles are subject to forward ($U \rightarrow M$; $\alpha$ in the figure, part a) and backward ($M \rightarrow U$; $\beta$ in the figure, part a) epimutation rates that are several orders of magnitude higher than DNA mutations, LD between SNP and DMR breaks down rapidly over time$^{31}$. At equilibrium ($t_\infty$), SNP alleles are expected to be completely disassociated from epialleles. The SNP is therefore no longer detected as meQTL, eQTL or phQTL (see the figure, right panel of part b).

Nonetheless, DMRs continue to cause differential gene expression (and affect complex traits) but now do so independently of the genotype of the flanking SNP. In this way, epigenetic variation can contribute to the heritability of complex traits without these contributions being captured by SNP-based genome-wide association scans$^{116}$. If epialleles affect fitness, selection can also shape epiallele frequencies at any time $t$ (not shown). At $t_\infty$, these frequencies are given by the selection–epimutation equilibrium$^{115,128,130,131,132}$. Epimutations with or without selection provide an evolutionary mechanism that can affect population epigenomic variation independently of genetic explanations. Recent population genetic models that account for forward–backward epimutations$^{115,130}$ can be used to test this hypothesis against empirical site-frequency spectra of DMRs or differentially methylated positions (DMPs). This approach provides a formal framework for genome-wide scans of epigenetic selection in natural populations.
transcription factor genes, which could be causal and explain this specificity. Although gene body methylation typically has no clear phenotype, it seems to be highly conserved on orthologues in various plant species and thus evolutionarily important, or the indirect result of an evolutionarily important process. Consistent with this observation, analysis of Swedish A. thaliana populations shows that GBM levels correlate with geographical and climatic variables, suggesting that they contribute to local adaptation. Indeed, northern Swedish accessions show increased GBM compared with southern accessions. This geographical divide is accompanied by allele-frequency shifts at most trans-acting meQTL, with ‘increasing’ alleles for GBM being more frequent in the north than in the south. Hence, epigenetic adaptation seems to be mediated, in this case, by selection on a large number of trans-acting meQTL, which is supported by the fact that many of these loci fall into regions of previously characterized selective sweeps.

Heritable epimutations may partly drive population epigenomic variation. Despite the detection of meQTL in both cis and trans, the two largest plant population epigenomic studies to date show that only 18–35% of all DMRs can be associated with genetic variation at genome-wide scale (Table 1). An intriguing hypothesis is that this lack of association is the result of the sequence-independent segregation of alternative methylation states (epialleles). In plants, heritable epialleles frequently arise de novo through germline epimutation events; that is, through stochastic losses or gains of DNA methylation. These heritable epimutations seem to occur mainly at CpG dinucleotides and are highly dependent on genomic context. Estimates in A. thaliana mutation accumulation lines indicate that the forward epimutation rate (that is, the rate of methylation gain) is about 2.56 × 10^{-4} and the backward epimutation rate (that is, the rate of methylation loss) is about 6.30 × 10^{-4} per CpG site, per haploid methylome, per generation. Because these rates are on average about five orders of magnitude higher than the known genetic mutation rate (~7 × 10^{-9} ) they provide one mechanism by which epigenetic variants can become disassociated from their underlying DNA sequence haplotypes over evolutionary timescales. The degree of disassociation depends on the precise epimutation rate, the age of the haplotype and the potential effect of epigenetic selection. Population epigenomic variation in plants could therefore be substantially shaped by epimutational processes. Although this biological hypothesis could certainly account for the modest proportion of genetic associations seen in genome-wide studies, it needs to be distinguished from more mundane technical explanations, such as low statistical power to detect meQTL, complex polygenic or epistatic genetic architecture, presence of causative rare alleles, and so on. These technical difficulties potentially undermine many ecological studies, particularly in non-model organisms, that report evidence of epigenetic adaptation without ruling out the possibility that such effects are mediated by selection on (undetected) cis- or trans-acting genetic variation.

Because of these technical issues, several groups have tried to assess the effects of epigenetic inheritance on population epigenomic variation in more simplified experimental systems in which confounding effects of genomic variation have been reduced to a minimum. Cortijo et al., for instance, showed that experimentally-induced DMRs in an isogenic A. thaliana population are remarkably stable and account for about 60% of the heritability of several plant complex traits. Interestingly, these experimentally-induced DMRs are also variable in natural populations of this species, suggesting that they are targets of epimutations in the wild and potentially also subject to natural selection. Observations such as these pose deeper questions about the evolutionary mechanisms that generate population epigenomic variation in plants and have stimulated substantial theoretical work in recent years.

Relating meQTL to chromatin states. Unlike in humans, population studies of integrated chromatin states have not been carried out in plants. It therefore remains unclear how the cis- and trans-genetic associations for DNA methylation manifest at the level of chromatin organization. DNA methylation interacts with several chromatin marks in plants. In A. thaliana DNA methylation is associated with the presence of H3K27me1, H3K9me2 and H4K20me1 and the absence of the H2A.Z histone variant, and the RNA-directed DNA methylation and CMT2–CMT3 methyltransferase pathways are dependent on the presence of H3K9me and the absence of H3K4me3. In humans, reference epigenomes have been instrumental to relate meQTL analysis back to chromatin state knowledge. Despite initial attempts to study reference epigenomes in plants, it therefore remains unclear how pre-existing epigenomic resources in plants and would be instrumental for dissecting the regulatory implications of meQTL, as well as for contextualizing genetic associations from genome-wide mapping studies of plant complex traits.

Conclusions
A detailed understanding of the genetic basis of population epigenomic variation is just beginning to emerge. Important quantities, such as the heritability of epigenomic variation at a locus, its distribution along the genome, or the effect sizes of cis- and trans-acting genetic variants, have so far only been partly studied in the context of DNA methylation, using relatively small sample sizes and tissue types. Although it may be premature to draw major conclusions from these studies, a trend is beginning to emerge.
One key observation from human methylation studies is that most functionally relevant changes in CpG methylation occur in regulatory elements, such as enhancers. These changes correlate with developmental transitions within an individual but seem also to be hubs of within- and between-population epigenomic variation. Popular array technologies do not measure these regions sufficiently well to be able to draw solid conclusions regarding genome-wide patterns of inter-individual differences and may therefore bias insights into the underlying genetic architecture. In light of recent methylome sequencing approaches, we hypothesize that array-based studies probably underestimate the prevalence of genetic contributions and may have led to downward biased heritability estimates in the literature. The severity of this bias will need to be determined as sequencing approaches become economically more feasible to be applied to large samples with sufficiently high sequencing depth.

 Nonetheless, most studies to date agree that inter-individual differences in DNA methylation are mainly determined by cis-regulatory sequence polymorphisms, probably involving mutations in TFBSs with downstream consequences on local chromatin environment. The sparsity of trans-acting polymorphisms in humans suggests that such effects are highly deleterious. Indeed, cis-acting factors are expected to be caused by mutations in chromatin control genes or other highly pleiotropic regulators. If trans-acting variants do exist in human populations, they probably segregate as rare alleles or originate from somatic mutations and present with clinical phenotypes, as is the case in many cancers.

 Trans effects on genome-wide DNA methylation states are much more common in plants. In fact, trans-acting variants are a major contributor to within- and between-population methylome variation and seem to contribute to local adaptation. The interpretation of cis associations in plants poses unique challenges. These challenges stem mainly from the fact that differential methylation states (epialleles) can be inherited meiotically for many generations, so that detected cis associations are simply a reflection of LD between sequence alleles and epialleles rather than the result of any type of regulatory interactions. These two possibilities cannot be confidently distinguished by genome-wide mapping analysis alone, and caution is therefore warranted for the interpretation of genetic associations. Unlike in mammalian systems in which population epigenomic studies are a tool to fill in the so-called genotype–phenotype map, researchers undertaking epigenomic analysis in plant populations face the challenge of having to delineate genetic from epigenetic inheritance, their respective contributions to regulatory variation and study how they co-evolve over evolutionary timescales.

 Beyond studies of DNA methylation, a more complete picture of how genetic variation influences chromatin state differences between individuals requires simultaneous measurements of multiple epigenetic marks. Such studies remain cost-prohibitive for large population samples. However, the first small-scale integrative studies, mostly in humans, show that single SNPs often affect multiple layers of chromatin organization as well as the expression of proximal and distal target genes through non-trivial regulatory mechanisms. We argue that these regulatory mechanisms could be further elucidated if genetic analyses were to treat 'combinatorial chromatin states' as molecular units of analysis, rather than individual epigenetic marks, as combinatorial chromatin states provide a more accurate description of the functional state of a locus. This approach will benefit tremendously from emerging single-cell or low-input epigenomic sequencing methods, because tissue-specific population epigenomic variation is more accurately captured and inferences of combinatorial chromatin states are less obscured by cellular heterogeneity.

 It is timely to scale up population genetic studies of chromatin states to include more diverse cell-types, epigenetic marks and individuals. Advances in sequencing technologies, computational algorithms, decreases in sequencing costs and a conceptual shift from individual marks to integrated chromatin states will be instrumental for elucidating the genetic sources of population epigenomic variation.


This human population study profiles DNA methylation in 100,000 individuals across the entire genome, and estimates that narrow-sense heritability of methylation levels at individual CpG sites is about 0.2 on average.


This is the largest, and arguably least biased, human association mapping study to date. The study used MB-MethylChip genotyping arrays that allow analysis of 697,567 methylomes of 697 individuals and identified many cis associations in enhancer and chromatin regions that are poorly covered by current array platforms.


This plant association mapping study integrates base-resolution methylation data with sequenced genomes of 152 Arabidopsis accessions.


This plant association mapping study of base-resolution methylation data with sequenced genomes of 152 Arabidopsis accessions.


This study analyzes multi-generational base-resolution methylation of A. thaliana mutation accumulation lines and provides the first estimates of forward and backward epimutation rates.

Epigenetic silencing of plant growth.

The epigenome and produces heritable changes in the response to selection.

The lifetime of monozygotic twins.

This study in Arabidopsis thaliana demonstrated that experimentally-induced epimutations can segregate for many generations and contribute to the heritability of plant complex traits.

Heritable epigenetic variation creates potential for evolution of plant phenotypic plasticity.

The authors declare no competing interests.

Acknowledgements

F.J. acknowledges support from the Technical University of Munich–Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement #291763. M.C.T. was supported by the Netherlands Organization for Scientific Research and by a University of Groningen Rosalind Franklin Fellowship.

Competing interests statement

The authors declare no competing interests.

FURTHER INFORMATION

ChromSig: http://bioinformatics.renlab.ucsd.edu/rentac/wiki/ChromSig

ChromHMM: https://chromhmm.org/ ChromHMM

chromstat: http://www.chromstat.org/hHMM: https://chromhmm.org/hiHMM

JUMOSAICs: https://bioinfo.donostia-santurtun.org/packages/release/bioinfo/jumosaic.html

Segway: https://www.epigenomics.ca/hiFiManallexi/ssegway

TreeHMM: https://github.com/uci-cbcl/tree-hmm

SUPPLEMENTARY INFORMATION

See online article: S1 (table) | S2 (figure) | S3 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF