◎ EPIGENETICS

Genetic sources of population epigenomic variation

Aaron Taudt[1], Maria Colomé-Tatché[1,2] and Frank Johannes[3,4]

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.

Chromatin state maps
Computationally integrated genome-wide measurements of different epigenetic marks.

Epigenome
The complete set of epigenetic marks at every genomic position in a given cell at a given time.

[1]*Quantitative Epigenetics, European Research Institute for the Biology of Ageing, University Medical Center Groningen, Antonius Deusinglaan 1, Groningen NL-9713 AV, The Netherlands.*
[2]*Institute for Computational Biology, Helmholtz Center Munich, Ingolstädter Landstrasse 1, Neuherberg 85764, Germany*
[3]*Population Epigenetics and Epigenomics, Technical University of Munich, Liesel-Beckmann-Strasse 2, Freising 85354, Germany.*
[4]*Institute for Advanced Study, Technical University of Munich, Lichtenbergstrasse 2a, Garching 85748, Germany.*

Correspondence to F.J.
frank@johanneslab.org

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 scale[1,2] (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 treatments[3–5]. 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 studies[5–7].

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 causes[8,9]. 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[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[1,2] (FIG. 1b).

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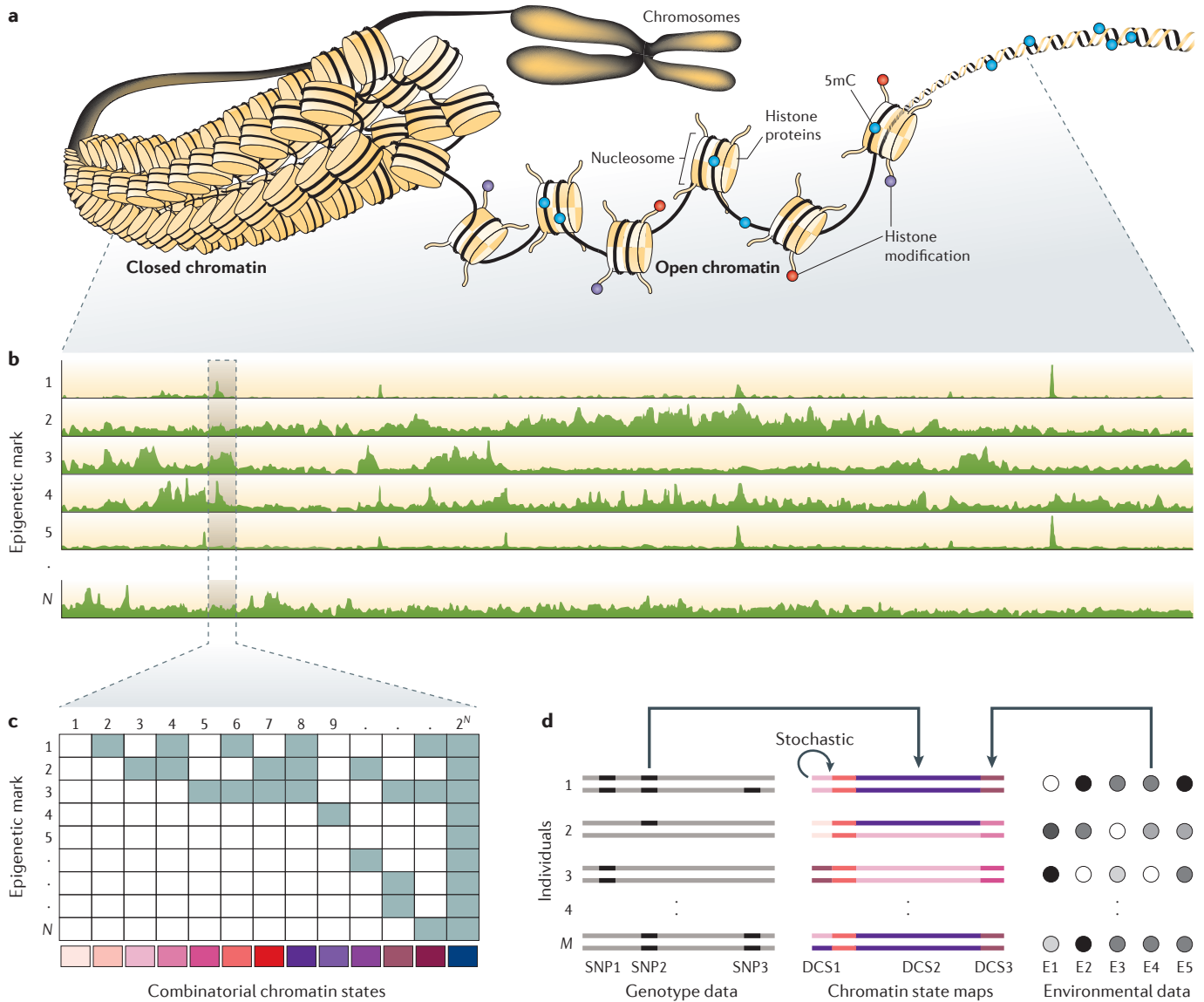

Figure 1 | **Main steps in population epigenomic analysis. a** | DNA is tightly packaged in cells and is functionally modified by a variety of epigenetic marks, such as cytosine methylation (5mC) or post-translational changes in histone proteins. The co-occurrence of specific epigenetic marks in a genomic 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 1 | **Overview of genetics studies of population epigenomic variation***

Organism	Population	Tissue	N ind.	N marks	Method	FDR control	cis (kb)	% assoc.	% cis	% trans	% var. cis	% var. trans	Refs
Integrated chromatin states													
Human	Natural	LCL	70	1‡‡	GWA-cis	0.1	20	0.64	100	NT	~30–40	NT	35
Human	Natural	LCL	14	5	AT	NR	NR	NR	100	NT	NR	NT	36
Human	Natural	LCL	19§	5	Other method	NR	NR	NR	100	NT	NR	NT	7
Human	Natural	LCL	10	4	GWA-cis	0.2	1	NR	100	NT	NR	NT	37
Human	Natural	LCL	64	5	GWA-cis	0.1	3	4.22	100	NT	NR	NT	42
Human	Natural	LCL	47	3	GWA-cis	0.1	250	8.1‖	100	NT	~40	NT	38
Human	Natural	LCL	75	3	GWA-cis	0.1	2¶	10–15#	100	20950**	NR	NR	39
Rat	RIL	Heart and liver	30	2	LM	0.05	10^4	NR	69.24, 75.28‡‡	30.76, 24.72‡‡	NR	NR	45
Yeast	RIL	NA	96	1‡	LM	0.01	100	NR	7.9	92.1	NR	33	47
Yeast	RIL	NA	94	1‡	LM	0.01	100	NR	15.7	84.3	NR	NR	48
DNA methylation in humans													
Human	Natural	CRBLM, FCTX, TCTX and pons	150	1	GWA	NR	10^3	4–5.1	30–40	60–70	18–88§§	18–88§§	65
Human	Natural	CRBLM	153	1	GWA	0.05	10^3	8.71‖‖	98	2	17–73	NR	66
Human	Natural	LCL	77	1	GWA	0.1	50	0.17	73	27	22–63	NR	67
Human	Natural	LCL	180§	1	GWA-cis	0.3–0.4	100	0.12–0.38	100	NT	36–92¶¶	NT	68
Human	Natural	WB	201	1	GWA-cis	NR	500	11.93‖‖	100	NT	NR	NT	69
Human	Natural	LCL	133§	1	GWA-cis	0.05	100	5.1–5.8	100	NT	30 ¶¶,##	NT	70
Human	Natural	FIB, T cells and LC	66–111	1	GWA-cis	0.1	5	3.4–7.8	100	NT	10–90	NT	71
Human	Natural	HPI	89	1	GWA	0.05	500	2.57	96.8	3.2	NR	NR	72
Human	Natural	FIB	62	1	GWA-cis	0.05	250	2‖‖	100	NT	NR	NT	73
Human	Natural	LCL	133§	1	GWA	0.05	10^3	1.8–2.6	100	NT	23–97¶¶	NT	74
Human	Natural	LCL	34§	1	GWA-cis	NR	200	NR	100	NT	26‖‖	NT	64
Human	Natural	WB	697	1	GWA	0.01	10^3	15	99.4	0.6	NR	NR	75
DNA methylation in plants													
Maize	Natural and RIL	Leaf	51	1	GWA-cis and LM	0.05	2	51	100	NT	NR	NT	83
A. thaliana	Natural and RIL	Leaf and MSIT	152	1	GWA	0.01	NR	16–35	26	74	NR	NR	81
Soybean	RIL	Leaf	83	1	LM	NR	CHR	91	97.5	2.5	51§§,##	51§§,##	84
A. thaliana	Natural	WR	155	1	GWA	0.05***	100	21‡‡‡, 18‖‖‖§§§	31‡‡‡, 45‖‖‖§§§	69‡‡‡, 55‖‖‖§§§	~10##,‡‡‡, ~20‖‖‖,##,§§§	~50##,‡‡‡, ~20‖‖‖,##,§§§	82

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. ‖Percentage of so-called variable chromatin 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 *cis*-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 code[13,14] 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 states[15,24,28,31,137]. 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 mark[16,138]. A third view defines probabilistic chromatin states[6,19,20,23,26] (also called fuzzy chromatin states[28]), 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.
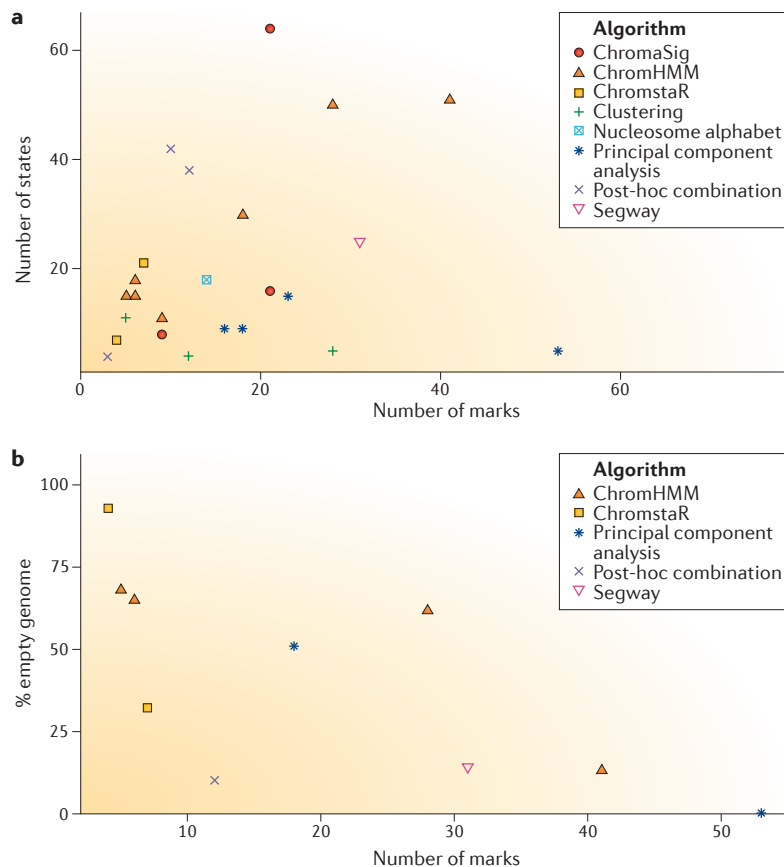
code is potentially complex, as with 100 marks there are $2^{100} \approx 1.3 \times 10^{30}$ possible combinations of modifications at any given nucleosome. Although at a mechanistic level there are chemical restrictions on the co-occurrence of certain marks, the measured signal is an average over different cells and convoluted with noise, and spurious combinations may therefore be detectable. Nonetheless, integrative analysis of genome-wide maps based on a subset of all histone modifications have so far consistently revealed that only a small proportion of all possible combinations exist in the epigenome[5–7,15–31] (BOX 1; Supplementary information S1 (table)). This fact hints at strong biological restrictions on the placement of epigenetic marks. Despite this reduction in complexity, the inference of integrative chromatin states from individual array or NGS measurements continues to pose major computational and conceptual challenges that have not been fully solved (BOX 1).

Chromatin states (BOX 1) define a language that efficiently summarizes information across different marks and enables comprehensive comparisons across tissues, developmental stages and individuals. Large-scale initiatives have made extensive use of those definitions and have produced reference epigenomes for various cell types and conditions in model and non-model species (Supplementary information S1 (table)). Comparisons of these reference epigenomes have provided several insights into epigenomic variation. A major insight is that chromatin states corresponding to enhancer elements are most variable between tissue types[5,20] and developmental time points[31,32], whereas chromatin signatures corresponding to transcribed regions, transcription start sites (TSSs) or repressed regions are less variable[5]. Certain elements termed cREDS (*cis*-regulatory elements with dynamic signatures) are found with a strong promoter signature in one tissue but with an enhancer signature in other tissue types[5,33], thus blurring the distinction between enhancer and promoter sequence elements[34].

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 humans[7,35–39] (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 programmes[40]. 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 (<5 kb) 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.*[7] 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, Polycomb-associated state H3K27me3 or empty state; see BOX 1), or vice versa. State transitions of this type reflect what is typically seen during cellular differentiation[31], suggesting that intra- and inter-individual epigenomic 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[41].

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[37–39,42]. 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[37]. Such coordinated, multi-layered chromatin changes can span several thousand base pairs and provide a mechanism by which hQTL act as expression quantitative trail loci (eQTL) to proximal genes without actually being located in coding regions[36,37].

Waszak *et al.*[38] and Grubert *et al.*[39] 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[43]. 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[44]; 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.*[45] 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. 46), 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[47,48]. *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 remodellers 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[7,36,49]. 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.

Histone quantitative trait loci
(hQTL). Genetic loci that contribute to variation in histone modification states and/or levels in *cis* or *trans*.

Expression quantitative trait loci
(eQTL). Genetic loci that contribute to variation in expression levels of mRNA in *cis* or *trans*.

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[50,51] with major roles in the regulation of gene expression and the silencing of transposable elements and repeat sequences[52,53]. In humans, the majority
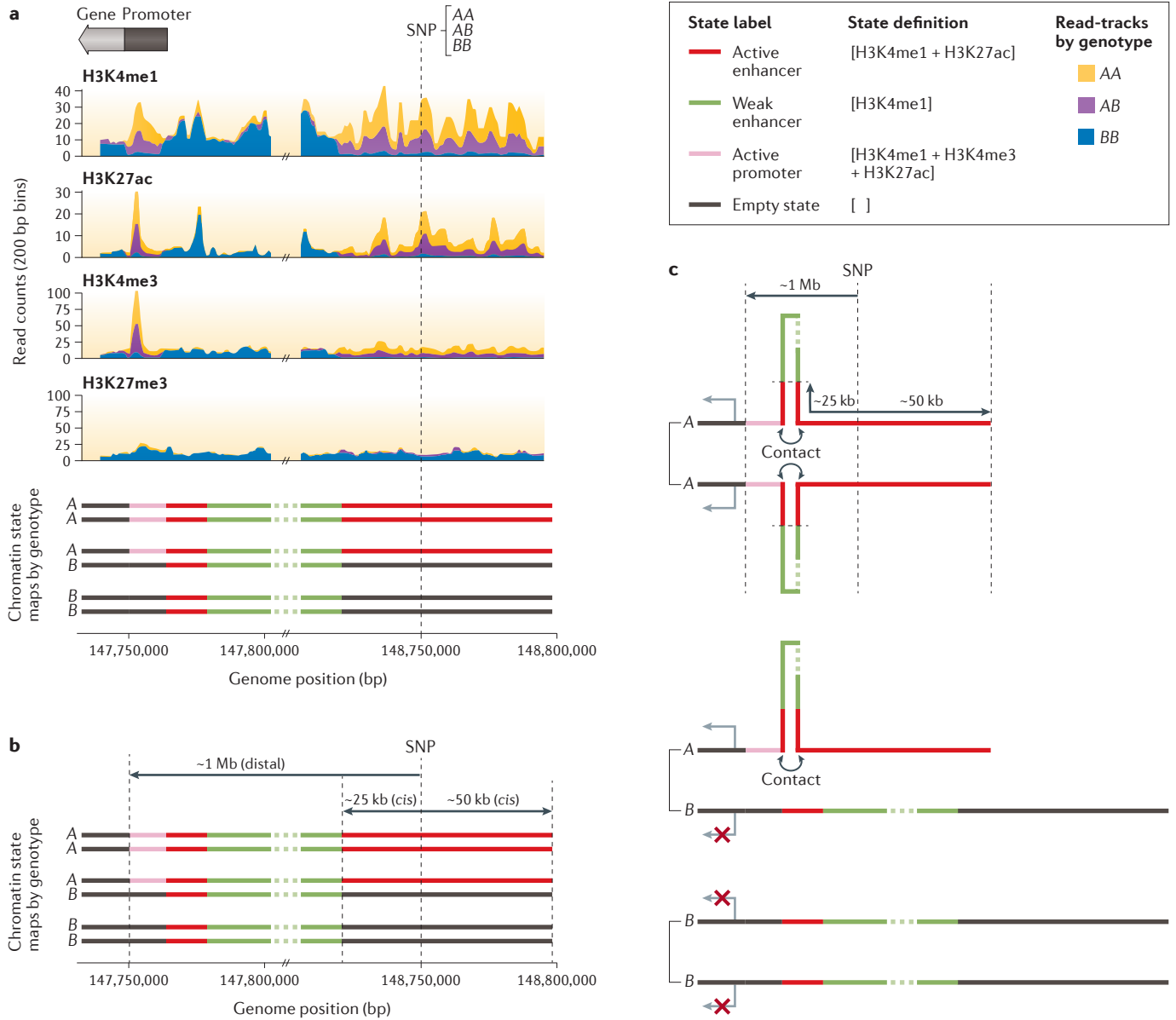

Figure 2 | **Single-nucleotide polymorphisms affecting chromatin states produce signatures of molecular pleiotropy. a** | Shown are the next-generation sequencing read-tracks of histone modifications: monomethylated histone H3 lysine 4 (H3K4me1), acetylated H3K27 (H3K27ac), trimethylated H3K4 (H3K4me3) 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[54,55]. These changes are coordinated at the level of chromatin states and involve substantial crosstalk with various histone modifications, such as H3K9me and H3K4me[56,57].

***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[58]. 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[59]. In one of the largest genetically informative studies to date, McRae *et al.*[60] 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[61,62]. Although this estimate may seem low, it is higher than the average heritability estimates obtained for gene expression levels[63], indicating that inter-individual differences in CpG methylation are under stronger genetic control[7,36,49,60,64].

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)[8]. 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' (REF. 139). Detection of such somatic epimutations will require advances in single-cell epigenomic sequencing technologies[140].

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[93]. Unlike in mammalian systems, germline 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.
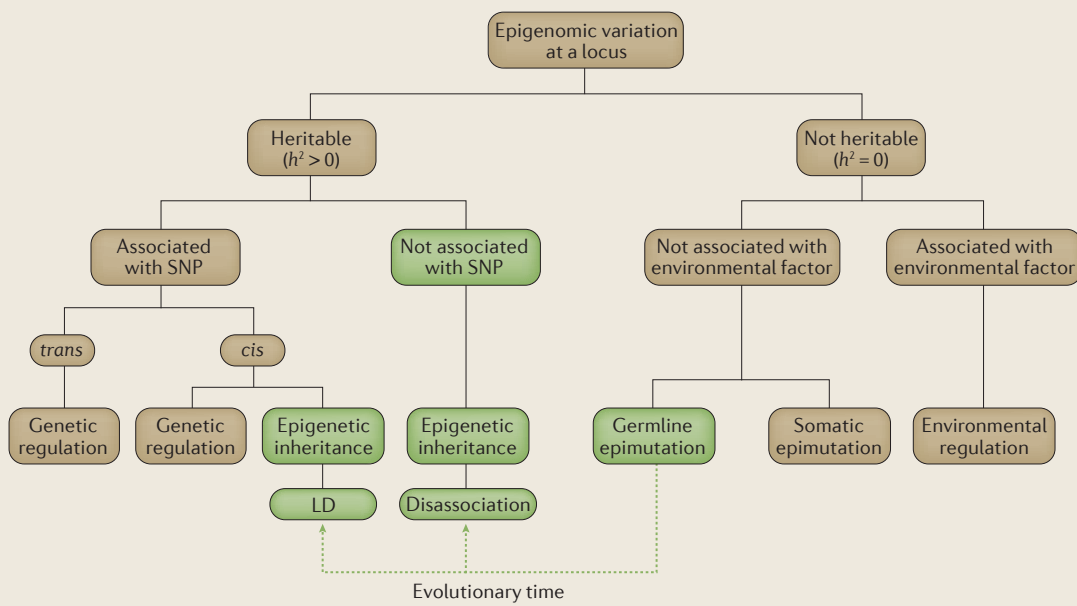
Several association mapping approaches have tried to identify specific *cis*- and *trans*-acting SNPs that account for the heritability in CpG methylation (that is, methylation QTL (meQTL))[42,64–75]. Despite major differences in cell types, data processing and analytical techniques, a remarkably consistent 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[42,65,71,72,74,75]. This mirrors the distribution of site-specific heritability estimates, which are relatively low in CpG islands and higher with increased distance from CpG islands[60]. 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[64,68,70,74], 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 islands[60,73,76], often in regions that are poorly surveyed by the 450k array[76,77], 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 methylome 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)[75], 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.*[42], 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%[35]) 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%[37]). 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[5]. 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.*[42] 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.[75] 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 thaliana*[81,82], maize[83] and soybean[84]. Although plant species differ widely in their total methylation content[85], most likely owing to differences in genome size and organization[85,86], patterns of intra-specific methylation variation seem to be broadly conserved[81–84,87–90]. Gene-rich euchromatic regions tend to be the most variable, whereas variation in transposable element (TE)-rich heterochromatic regions is largely suppressed[81,83,91–93]. The lack of variation in heterochromatic regions is consistent with robust silencing of TE sequences by small-RNA-directed mechanisms[52]. Unlike mammalian genomes in which methylation differences at single CpG sites can have functional consequences[80], 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.*[81] 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 maize[83]. 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 100 kb). Slightly more prevalent cis effects (31–45% of all associations) were reported by Dubin *et al.*[82], 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 maize[83] and RILs of

soybean[84]. 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 sequences[94,95]. 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 changes[96,97]. Interestingly, the spreading of DNA methylation from structural variant alleles seems to be partly stochastic and thus varies between individuals both in extent and stability[94,95]. 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, respectively[81,83]. 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-wide[98,99]. 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 meQTL[82,91,100]. An instructive example comes from the mapping

---

Differentially methylated regions
(DMRs). Regions of DNA that have different methylation patterns between samples or individuals.

Epialleles
Alternative chromatin states at a given locus. Typically, they refer to alternative DNA methylation states, although in principle they could also refer to changes in other epigenetic marks.

Gene body methylation
(GBM). Average CG methylation levels over the body of genes.

analysis of CHH methylation (where H can be any base except G) in local populations of Swedish *A. thaliana* accessions. Dubin *et al.*[82] found that two loss-of-function SNPs in a single gene, CHROMOMETHYLASE 2 (*CMT2*), accounted for 23% of all detected *trans* effects (>100 kb). *CMT2* encodes a homologue of the methyltransferase CMT3 (REF. 101), which interacts with H3K9me to catalyse CHG and CHH methylation in heterochromatin-associated long TEs[98]. Interestingly, CMT2 has negligible *trans* effects on CHH methylation levels in the worldwide accessions[100], because the causative *CMT2* alleles are either not present or occur

at very low frequencies[82]. This observation suggests that epigenomic variation among plant populations can vary substantially on the basis of allele frequency differences at a few crucial chromatin-control genes.

Extensive *trans* effects have also been reported for gene body methylation (GBM) levels[82] or for DMRs overlapping genic sequences[91], which account for 55–70% of all detected associations. Unlike in the case of *CMT2*, however, meQTL that affect GBM in *trans* seem to be much less pleiotropic, often only affecting a handful of target sequences. Regions in LD with these meQTL are enriched for transcription regulators, such as

---

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_n (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 ($U \rightarrow M$; α in the figure, part **a**) and backward ($M \rightarrow U$; β 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[93]. At equilibrium (t_∞), 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[106]. If epialleles affect fitness, selection can also shape epiallele frequencies at any time *t* (not shown). At t_∞, these frequencies are given by the selection–epimutation equilibrium[93,128,130,132]. 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[130,132] 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.
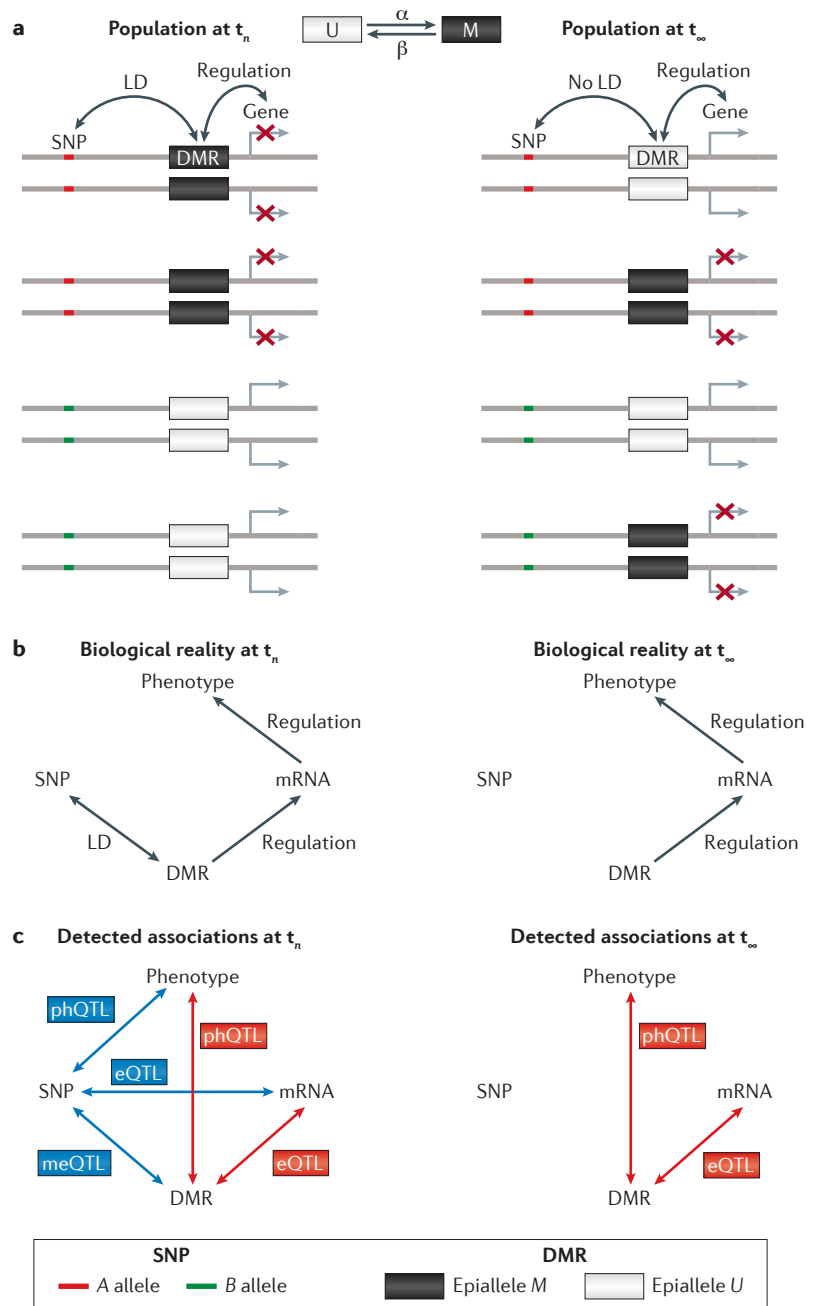
transcription factor genes, which could be causal and explain this specificity. Although gene body methylation typically has no clear phenotype[94], it seems to be highly conserved on orthologues in various plant species and thus evolutionarily important[102–104], or the indirect result of an evolutionarily 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[105].

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)[106]. 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[93,107–109]. 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[93]. Because these rates are on average about five orders of magnitude higher than the known genetic mutation rate ($\sim 7 \times 10^{-9}$) (REF. 110) they provide one mechanism by which epigenetic variants can become disassociated from their underlying DNA sequence haplotypes over evolutionary timescales[93] (BOXES 2,3). 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 architectures, 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[111–122]. Cortijo *et al.*[117], 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[123–132]. It is precisely the transgenerational dimension in plant population epigenomics that makes it fundamentally different, and arguably more challenging, than population epigenomics in other organisms in which epigenetic inheritance is negligible.

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 (REF. 24) and the absence of the H2A.Z histone variant[133], and the RNA-directed DNA methylation and CMT2–CMT3 methylation maintenance pathways are dependent on the presence of H3K9me[57,134,135] and the absence of H3K4me3 (REF. 57). 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[15,24,29,30], no such large-scale integrated reference epigenomes are currently available. Recently, Lane *et al.*[136] called for the launch of a plant ENCODE (Encyclopedia of DNA Elements) project. This project would benefit from the large 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.

Epimutation
Heritable stochastic change in chromatin state at a given position or region. In the context of cytosine methylation, epimutations are defined as heritable stochastic changes in the methylation status of a single cytosine or of a region or cluster of cytosines. Such changes do not necessarily imply changes in gene expression.

REVIEWS

Genotype–phenotype map
A map that describes the functional connections between genotype and phenotype. The concept of a genotype–phenotype map is widely used as a metaphor for the many ways in which genotypic information influences the phenotype of an organism.

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, *trans*-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.

1. Park, P. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669–680 (2009).
2. Laird, P. W. Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.* **11**, 191–203 (2010).
3. Adams, D. *et al.* BLUEPRINT to decode the epigenetic signature written in blood. *Nat. Biotechnol.* **30**, 224–226 (2012).
4. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
5. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
6. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
7. Kasowski, M. *et al.* Extensive variation in chromatin states across humans. *Science* **342**, 750–752 (2013).
   This human population genetic study is the only study to date to use chromatin states as molecular units of analysis to assess epigenomic differences between individuals.
8. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12**, 529–541 (2011).
9. Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. *Nat. Rev. Genet.* **14**, 585–594 (2013).
10. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).
11. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
12. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
13. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
14. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
15. Li, X. *et al.* High-resolution mapping of epigenetic modifications of the rice genome uncovers interplay between DNA methylation, histone methylation, and gene expression. *Plant Cell* **20**, 259–276 (2008).
16. Hon, G., Ren, B. & Wang, W. ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Comput. Biol.* **4**, e1000201 (2008).
17. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
18. Hon, G., Wang, W. & Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* **5**, e1000566 (2009).
19. Roy, S. *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
20. Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480–485 (2011).
21. Riddle, N. C. *et al.* Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res.* **21**, 147–163 (2010).
22. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
23. Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).

24. Roudier, F. et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J. 30, 1928–1938 (2011).

25. Liu, T. et al. Broad chromosomal domains of histone modification patterns in C. elegans. Genome Res. 21, 227–236 (2011).

26. Hoffman, M. M. et al. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nat. Methods 9, 473–476 (2012).

27. Lai, W. K. M. & Buck, M. J. An integrative approach to understanding the combinatorial histone code at functional elements. Bioinformatics 29, 2231–2237 (2013).

28. Luo, C. et al. Integrative analysis of chromatin states in Arabidopsis identified potential regulatory mechanisms for natural antisense transcript production. Plant J. 73, 77–90 (2013).

29. Sequeira-Mendes, J. et al. The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. Plant Cell 26, 2351–2366 (2014).

30. Baker, K. et al. Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. Plant J. 84, 111–124 (2015).

31. Taudt, A. Nguyen, M. A., Heinig, M., Johannes, F. & Colome-Tatche, M. chromstaR: Tracking combinatorial chromatin state dynamics in space and time. bioRxiv http://dx.doi.org/10.1101/038612 (2016).

32. Lara-Astiaso, D. et al. Chromatin state dynamics during blood formation. Science 55, 1–10 (2014).

33. Leung, D. et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature 518, 350–354 (2015).

34. Andersson, R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. Bioessays 37, 314–323 (2015).

35. Degner, J. F. et al. DNaseI sensitivity QTLs are a major determinant of human expression variation. Nature 482, 390–394 (2012).

36. Kilpinen, H. et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science 342, 744–747 (2013).

37. McVicker, G. et al. Identification of genetic variants that affect histone modifications in human cells. Science 342, 747–749 (2013).

38. Waszak, S. et al. Population variation and genetic control of modular chromatin architecture in humans. Cell 162, 1039–1050 (2015).
This human association mapping study analyses 3 histone modifications from 47 sequenced individuals and shows that cis-regulatory elements are preferentially found within the same chromosomal contact domain.

39. Grubert, F. et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell 162, 1051–1065 (2015).
This human association mapping study integrates chromatin profiles of 3 histone modifications from 75 sequenced individuals with Hi-C and chromatin interaction analysis by paired-end sequencing (ChIA-PET)-based chromatin contact maps and finds that distal hQTL are enriched within topologically associated domains.

40. Kimura, H. Histone modifications for human epigenome analysis. J. Hum. Genet. 58, 439–445 (2013).

41. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331–336 (2015).

42. Banovich, N. E. et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. 10, e1004663 (2014).
This human association mapping study is the only study to date to integrate methylation and histone modification data collected on the same individuals.

43. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480, 490–495 (2011).

44. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat. Rev. Genet. 8, 286–298 (2007).

45. Rintisch, C. et al. Natural variation of histone modification and its impact on gene expression in the rat genome. Genome Res. 24, 942–953 (2014).

46. Wysocka, J. et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell 121, 859–872 (2005).

47. Lee, K. et al. Genetic landscape of open chromatin in yeast. PLoS Genet. 9, e1003229 (2013).

48. Chai, X., Nagarajan, S., Kim, K., Lee, K. & Choi, J. K. Regulation of the boundaries of accessible chromatin. PLoS Genet. 9, e1003778 (2013).

49. McDaniell, R. et al. Heritable individual-specific and allele-specific chromatin signatures in humans. Science 328, 235–239 (2010).

50. Feng, S. et al. Conservation and divergence of methylation patterning in plants and animals. Proc. Natl Acad. Sci. USA 107, 8689–8694 (2010).

51. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916–919 (2010).

52. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11, 204–220 (2010).

53. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14, 204–220 (2013).

54. Ziller, M. J. et al. Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477–481 (2013).

55. Gifford, C. A. et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell 153, 1149–1163 (2013).

56. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. 10, 295–304 (2009).

57. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell. Biol. 16, 519–532 (2015).

58. Michels, K. B. et al. Recommendations for the design and analysis of epigenome-wide association studies. Nat. Methods 10, 949–955 (2013).

59. Dedeurwaerder, S. et al. Evaluation of the Infinium Methylation 450K technology. Epigenomics 3, 771–784 (2011).

60. McRae, A. F. et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol. 15, R73 (2014).
This human population study profiles DNA methylation in 614 individuals with the 450k array, and estimates that narrow-sense heritability of methylation levels at individual CpG sites is about 0.2 on average.

61. Bell, J. T. et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. 8, e1002629 (2012).

62. Gordon, L. et al. Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. Genome Res. 22, 1395–1406 (2012).

63. Wright, F. A. et al. Heritability and genomics of gene expression in peripheral blood. Nat. Genet. 46, 430–437 (2014).

64. Carja, O. et al. Worldwide patterns of human epigenetic variation. bioRxiv http://dx.doi.org/10.1101/021931 (2015).

65. Gibbs, J. R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. 6, e1000952 (2010).

66. Zhang, D. et al. Genetic control of individual differences in gene-specific methylation in human brain. Am. J. Hum. Genet. 86, 411–419 (2010).

67. Bell, J. T. et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 12, R10 (2011).

68. Fraser, H. B., Lam, L. L., Neumann, S. M. & Kobor, M. S. Population-specificity of human DNA methylation. Genome Biol. 13, R8 (2012).

69. van Eijk, K. R. et al. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. BMC Genomics 13, 636 (2012).

70. Moen, E. L. et al. Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. Genetics 194, 987–996 (2013).

71. Gutierrez-Arcelus, M. et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife 2, e00523 (2013).

72. Olsson, A. H. et al. Genome-wide associations between genetic and epigenetic variation influence mRNA expression and insulin secretion in human pancreatic islets. PLoS Genet. 10, e1004735 (2014).

73. Wagner, J. R. et al. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. 15, R37 (2014).

74. Zhang, X. et al. Linking the genetic architecture of cytosine modifications with human complex traits. Hum. Mol. Genet. 23, 5893–5905 (2014).

75. McClay, J. L. et al. High density methylation QTL analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. Genome Biol. 16, 291 (2015).
This is the largest, and arguably least biased, human association mapping study to date. The study used MBD-seq to determine the blood methylomes of 697 individuals and identified many cis associations in enhancer and heterochromatic regions that are poorly covered by current array platforms.

76. Busche, S. et al. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. Genome Biol. 16, 290 (2015).

77. Morris, T. J. & Beck, S. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. Methods 72, 3–8 (2015).

78. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics (2015).

79. Shoemaker, R., Deng, J., Wang, W. & Zhang, K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res. 20, 883–889 (2010).

80. Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature 528, 575–579 (2015).

81. Schmitz, R. J. et al. Patterns of population epigenomic diversity. Nature 495, 193–198 (2013).
This plant association mapping study integrates base-resolution methylomes with fully sequenced genomes of 152 A. thaliana accessions.

82. Dubin, M. J. et al. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. eLife 4, e05255 (2015).
This plant association mapping study of base-resolution methylomes provides evidence that selection on trans-acting meQTL for gene body methylation contribute to local adaptation.

83. Eichten, S. R. et al. Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell 25, 2783–2797 (2013).

84. Schmitz, R. J. et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Res. 23, 1663–1674 (2013).

85. Alonso, C., Perez, R., Bazaga, P. & Herrera, C. M. Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. Front. Genet. 6, 4 (2015).

86. Seymour, D. K., Koenig, D., Hagmann, J., Becker, C. & Weigel, D. Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. PLoS Genet. 10, e1004785 (2014).

87. Regulski, M. et al. The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res. 23, 1651–1662 (2013).

88. Vaughn, M. W. et al. Epigenetic natural variation in Arabidopsis thaliana. PLoS Biol. 5, 1617–1629 (2007).

89. Shen, H. et al. Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. Plant Cell 24, 875–892 (2012).

90. Hagmann, J. et al. Century-scale methylome stability in a recently diverged Arabidopsis thaliana lineage. PLoS Genet. 11, e1004920 (2015).

91. Schmitz, R. J. & Ecker, J. R. Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends Plant Sci. 17, 149–154 (2012).

92. Becker, C. & Weigel, D. Epigenetic variation: origin and transgenerational inheritance. Curr. Opin. Plant Biol. 15, 562–567 (2012).

93. van der Graaf, A. et al. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. Proc. Natl Acad. Sci. USA 6676–6681 (2015).
This study analyses multi-generational base-resolution methylomes of A. thaliana mutation accumulation lines and provides the first estimates of forward and backward epimutation rates.

94. Pecinka, A., Abdelsamad, A. & Vu, G. T. H. Hidden genetic nature of epigenetic natural variation in plants. Trends Plant Sci. 18, 625–632 (2013).

95. Fujimoto, R. *et al.* Molecular mechanisms of epigenetic variation in plants. *Int. J. Mol. Sci.* **13**, 9900–9922 (2012).

96. Hollister, J. D. & Gaut, B. S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* **19**, 1419–1428 (2009).

97. Bennetzen, J. L. & Wang, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530 (2014).

98. Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V. & Jacobsen, S. E. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).

99. Li, Q. *et al.* Genetic perturbation of the maize methylome. *Plant Cell* **26**, 4602–4616 (2014).

100. Shen, X. *et al.* Natural *CMT2* variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet.* **10**, e1004842 (2014).

101. Lindroth, A. M. *et al.* Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. *Science* **292**, 2077–2080 (2001).

102. Takuno, S. & Gaut, B. S. Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Mol. Biol. Evol.* **29**, 219–227 (2012).

103. Takuno, S. & Gaut, B. S. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc. Natl Acad. Sci. USA* **110**, 1797–1802 (2013).

104. Takuno, S., Ran, J.-H. & Gaut, B. S. Evolutionary patterns of genic DNA methylation vary across land plants. *Nat. Plants* **2**, 15222 (2016).

105. Long, Q. *et al.* Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* **45**, 884–890 (2013).

106. Johannes, F., Colot, V. & Jansen, R. C. Epigenome dynamics: a quantitative genetics perspective. *Nat. Rev. Genet.* **9**, 883–890 (2008).

107. Becker, C. *et al.* Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**, 245–249 (2011).

108. Schmitz, R. J. *et al.* Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**, 369–373 (2011).

109. Jiang, C. *et al.* Environmentally responsive genome-wide accumulation of *de novo Arabidopsis thaliana* mutations and epimutations. *Genome Res.* **24**, 1821–1829 (2014).

110. Ossowski, S. *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010).

111. Virdi, K. S. *et al. Arabidopsis MSH1* mutation alters the epigenome and produces heritable changes in plant growth. *Nat. Commun.* **6**, 6386 (2015).

112. Eichten, S. R. *et al.* Heritable epigenetic variation among maize inbreds. *PLoS Genet.* **7**, e1002372 (2011).

113. Amoah, S. *et al.* A hypomethylated population of *Brassica rapa* for forward and reverse epi-genetics. *BMC Plant Biol.* **12**, 193 (2012).

114. Hauben, M. *et al.* Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *Proc. Natl Acad. Sci. USA* **106**, 20109–20114 (2009).

115. Bossdorf, O., Arcuri, D., Richards, C. L. & Pigliucci, M. Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in *Arabidopsis thaliana*. *Evol. Ecol.* **24**, 541–553 (2010).

116. Li, Q., Eichten, S. R., Hermanson, P. J. & Springer, N. M. Inheritance patterns and stability of DNA methylation variation in maize near-isogenic lines. *Genetics* **196**, 667–676 (2014).

117. Cortijo, S. *et al.* Mapping the epigenetic basis of complex traits. *Science* **343**, 1145–1148 (2014). **This study in *A. thaliana* demonstrates that experimentally-induced epimutations can segregate for many generations and contribute to the heritability of plant complex traits.**

118. Verhoeven, K. J. F., Jansen, J. J., van Dijk, P. J. & Biere, A. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* **185**, 1108–1118 (2010).

119. Reinders, J. *et al.* Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* **23**, 939–950 (2009).

120. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**, e1000530 (2009).

121. Roux, F. *et al.* Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics* **188**, 1015–1017 (2011).

122. Zhang, Y.-Y., Fischer, M., Colot, V. & Bossdorf, O. Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytol.* **197**, 314–322 (2013).

123. Slatkin, M. Epigenetic inheritance and the missing heritability problem. *Genetics* **182**, 845–850 (2009).

124. Johannes, F. & Colome-Tatche, M. Quantitative epigenetics through epigenomic perturbation of isogenic lines. *Genetics* **188**, 215–227 (2011).

125. Day, T. & Bonduriansky, R. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. *Am. Nat.* **178**, E18– E36 (2011).

126. Geoghegan, J. L. & Spencer, H. G. Population-epigenetic models of selection. *Theor. Popul. Biol.* **81**, 232–242 (2012).

127. Klironomos, F. D., Berg, J. & Collins, S. How epigenetic mutations can affect genetic evolution: model and mechanism. *Bioessays* **35**, 571–578 (2013).

128. Furrow, R. E. Epigenetic inheritance, epimutation, and the response to selection. *PLoS ONE* **9**, e101559 (2014).

129. Furrow, R. E. & Feldman, M. W. Genetic variation and the evolution of epigenetic regulation. *Evolution* **68**, 673–683 (2014).

130. Charlesworth, B. & Jain, K. Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. *Genetics* **198**, 1587–1602 (2014).

131. Kronholm, I. & Collins, S. Epigenetic mutations can both help and hinder adaptive evolution. *Mol. Ecol.* **25**, 1856–1868 (2015).

132. Wang, J. & Fan, C. A neutrality test for detecting selection on DNA methylation using single methylation polymorphism frequency spectrum. *Genome Biol. Evol.* **7**, 154–171 (2015).

133. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129 (2008).

134. Jackson, J. P. *et al.* Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma* **112**, 308–315 (2004).

135. Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M. & Jacobsen, S. E. Genomewide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* **3**, e3156 (2008).

136. Lane, A. K., Niederhuth, C. E., Ji, L. & Schmitz, R. J. pENCODE: a plant encyclopedia of DNA elements. *Annu. Rev. Genet.* **48**, 49–70 (2014).

137. Zeng, X. *et al.* jMOSAiCS: joint analysis of multiple ChIP–seq datasets. *Genome Biol.* **14**, R38 (2013).

138. Won, K.-J. *et al.* Comparative annotation of functional regions in the human genome using epigenomic data. *Nucleic Acids Res.* **41**, 4423–4432 (2013).

139. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102**, 10604–10609 (2005).

140. Schwartzman, O. & Tanay, A. Single-cell epigenomics: techniques and emerging applications. *Nat. Rev. Genet.* **16**, 716–726 (2015).

FURTHER INFORMATION

ChromaSig: http://bioinformatics-renlab.ucsd.edu/rentrac/wiki/ChromaSig
ChromHMM: http://compbio.mit.edu/ChromHMM
chromstaR: http://www.chromstar.org
hiHMM: https://github.com/kasohn/hiHMM
jMOSAiCS: https://www.bioconductor.org/packages/release/bioc/html/jmosaics.html
Segway: https://www.pmgenomics.ca/hoffmanlab/proj/segway
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