

Capturing functional epigenomes for insight into metabolic diseases

Fiona Allum^{1,2}, Elin Grundberg^{3,*}

ABSTRACT

Background: Metabolic diseases such as obesity are known to be driven by both environmental and genetic factors. Although genome-wide association studies of common variants and their impact on complex traits have provided some biological insight into disease etiology, identified genetic variants have been found to contribute only a small proportion to disease heritability, and to map mainly to non-coding regions of the genome. To link variants to function, association studies of cellular traits, such as epigenetic marks, in disease-relevant tissues are commonly applied.

Scope of the review: We review large-scale efforts to generate genome-wide maps of coordinated epigenetic marks and their utility in complex disease dissection with a focus on DNA methylation. We contrast DNA methylation profiling methods and discuss the advantages of using targeted methods for single-base resolution assessments of methylation levels across tissue-specific regulatory regions to deepen our understanding of contributing factors leading to complex diseases.

Major conclusions: Large-scale assessments of DNA methylation patterns in metabolic disease-linked study cohorts have provided insight into the impact of variable epigenetic variants in disease etiology. In-depth profiling of epigenetic marks at regulatory regions, particularly at tissue-specific elements, will be key to dissect the genetic and environmental components contributing to metabolic disease onset and progression.

© 2020 The Author(s). Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords Metabolic diseases; Epigenomics; Regulatory elements; DNA methylation; Next-generation sequencing; Adipose tissue

1. INTRODUCTION

Metabolic diseases such as obesity are known to be multifactorial in nature, with genetic and environmental factors underlying disease onset (Figure 1) [1–4]. Case-control-based genetic studies of clinically ascertained obesity status have returned few cohesive results [5], most likely due to the heterogeneity of this condition. Conversely, quantitative genetic studies of complex traits related to metabolic diseases have successfully hinted at essential biological pathways involved in trait etiology; for instance, the central nervous system in genome-wide association studies (GWAS) of body mass index (BMI) [5–8] and adipogenesis in GWAS of waist-to-hip ratio [5,9]—a measure of abdominal adiposity. However, disease and trait-linked common genetic variants (i.e., single nucleotide polymorphisms; SNPs), per se, have been assessed to contribute only a limited proportion to the overall trait variance. Tabulations of total contributions vary from study to study and continue to improve with cohort sizes and the advent of statistical methods. For instance, the predicted genetic contribution to the obesity-related trait of BMI was initially assessed from twin studies to be between 40 and 70% [10–12], but alternative statistical methods have suggested that these numbers may be overestimated, and rather range from 30% to 40% [13]. Nevertheless, most reports have estimated that identified GWAS SNPs cumulatively account for less than 5–15% of observed variance, with gene–gene and gene–

environment effects and lack of coverage of rare variants suggested to contribute to the underestimation of identified trait-linked loci [14].

In addition, the majority of disease and trait-linked SNPs have been shown to locate to the non-coding portion of the genome and, more specifically, to the active chromatin in cells linked to the disease under study [15–19]—where the target gene affected is usually unknown. As such, findings from GWAS alone have not been easily translated into new biological mechanisms. Integrational approaches that layer additional association studies of cellular traits are required to functionally annotate these trait-linked SNPs. For this purpose, genome-wide gene expression assessments in genotyped tissue or cell samples from populations remain a common approach for functional analysis of disease SNPs and are referred to as expression quantitative trait loci (eQTL) analysis. Large-scale efforts to generate eQTL maps across multiple tissues — such as those by the Genotype-Tissue Expression (GTEx) consortium — have showed that SNP-expression associations have a strong tissue- and cell-specific component. As such, the results from these studies have pointed toward the need for careful consideration in tissue selection for the integrational functional analysis of disease SNPs to gain additional insight into disease etiology [20–25].

Annotation efforts have more recently been extended to expression-modulating epigenetic traits such as DNA methylation (5' methylcytosine) levels (methylation QTL; metQTL) or chromatin mark peaks (histone QTL; hQTL) with the same purpose as eQTL studies, which is

¹Department of Human Genetics, McGill University, Montréal, Québec, H3A 0C7, Canada ²McGill University and Genome Quebec Innovation Centre, Montréal, Québec, H3A 0G1, Canada ³Children's Mercy Kansas City, Kansas City, MO, 64108, United States

*Corresponding author. E-mail: egrundberg@cmh.edu (E. Grundberg).

Received September 27, 2019 • Revision received December 23, 2019 • Accepted December 30, 2019 • Available online xxx

<https://doi.org/10.1016/j.molmet.2019.12.016>

Review

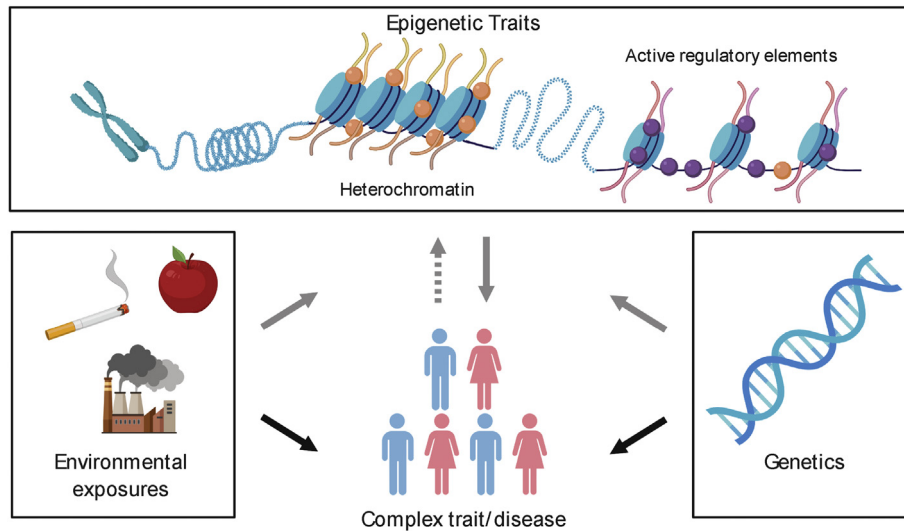


Figure 1: Summary of complex diseases and trait drivers. Graphical representation of contributing factors underlying complex disease risk including the interplay between environmental and genetic factors (black arrows) that impact epigenetics traits and provide insight into molecular mechanisms underlying disease etiology (gray arrows). This potential “causal pathway” (solid arrows) involving epigenetic changes impacting phenotype and disease is distinguished from a “reactive pathway”, in which the phenotype or disease state causes an epigenetic change (dotted arrow). Epigenetic traits are referred to here by CpG (DNA) methylation and histone modifications. Methylated CpGs are shown in orange and unmethylated CpGs are depicted in purple. Similarly, histone tails enriched in modifications correlating to inactive chromatin are shown in yellow (i.e., H3K9me2/3 and H3K27me3) and those enriched at active regulatory elements are presented in pink (i.e., H3K4me1, H3K4me3, and H3K27ac). Figure was created with [Biorender.com](https://www.biorender.com).

to inform about functional consequences of disease SNPs. Epigenetic traits are not only known to have similarly cell-specific signatures, but are also associated with stable long-term alterations caused by both genetic and environmental factors (Figure 1). In this way, genome-wide assessment of epigenetic traits, often referred to as epigenome mapping, can be used to gain biological insight into the disease model by enabling us to correlate genetics and environment to disease-linked phenotypes. With the tight link between epigenetic marks and gene activity regulation, epigenome mapping focuses on regulatory element profiling. Here, the proximal promoter regions can be more easily annotated as they usually contain the transcription start site (TSS) of genes. On the other hand, distal regulatory regions (i.e., enhancers) are more difficult to identify. These elements serve to modify the transcription rate of genes (proximal and/or non-proximal), which they physically contact through chromatin looping secured by protein interactions. Genome-wide identification of these regions poses challenges in the field, as gene expression is dependent on the differentiated nature of cells and is spatially and temporally modulated. Here, we aim to provide a review of the current literature on accessing and assessing high-resolution epigenetic signatures, mainly DNA methylation, with a specific focus on metabolic disease studies. We emphasize the benefit of using integrational studies by layering available epigenomic maps from large-scale consortium efforts, as well as recent efforts investigating regulatory regions in tissues and cells linked to the disease model for increased discovery power of novel disease genetic and epigenetic variants.

2. CONSORTIA EFFORTS TO DEFINE EPIGENOME MAPS

The Human Genome Project (HGP) was a ground-breaking collaborative effort that paved the way for subsequent large-scale projects [26,27] by providing the scientific community with the first map of the human genome. With the realization that the human genome was composed to a large extent of non-coding DNA for which functional interpretation was lacking, functional genomics studies gained attention to assign genomic

roles and identify regulatory elements [26]. To this end, the National Human Genome Research Institute (NHGRI) launched The Encyclopedia of DNA Elements (ENCODE) project in 2003, with the purpose of providing detailed maps of regulatory elements across the human genome [28,29]. With the ENCODE project focusing on cultured cell lines for genome characterization, the subsequent National Institutes of Health (NIH) Roadmap Consortium shifted toward providing reference epigenome mapping in *ex-vivo* tissues and primary cells [30]. In 2015, these efforts culminated in the public release and integrative analysis of over 100 reference human epigenomes [16]. These reference epigenome efforts were unified with similar consortia from North America, Europe, Australia, and Asia to form The International Human Epigenome Consortium (IHEC), with the goal of providing access to over 1000 complete human epigenome profiles [31]. One focus area of the consortium is the analysis of cells connected to metabolic diseases, such as obesity and steatosis. As of the latest IHEC portal (<http://epigenomes.ca/ihec>) release (2018-10; hg19 reference), there are 824 metabolic disease-relevant epigenomes derived from various adipose depots, liver, muscle, gastrointestinal, and brain tissues, generated from six consortia efforts (Figure 2). Other more-targeted studies have also contributed valuable reference and population epigenome maps, including DNA methylation profiles of adipocytes [32], subcutaneous and visceral adipose tissues [33–35], open chromatin maps of adipocytes [34], and transcriptome profiles of adipocytes [32,34], subcutaneous adipose tissues [21], and liver tissues [36]. Together, these efforts provide a rich resource of epigenomes to the scientific community to help decipher the genetic and environmental basis of disease and regulation of gene expression, as well as advance technological and statistical methods to discover and interpret epigenomic marks.

3. EPIGENOME ASSAYS IN POPULATION- AND DISEASE-BASED STUDIES

Assessing genome-wide epigenetic patterns generated by the consortia has provided us with a wealth of information concerning the

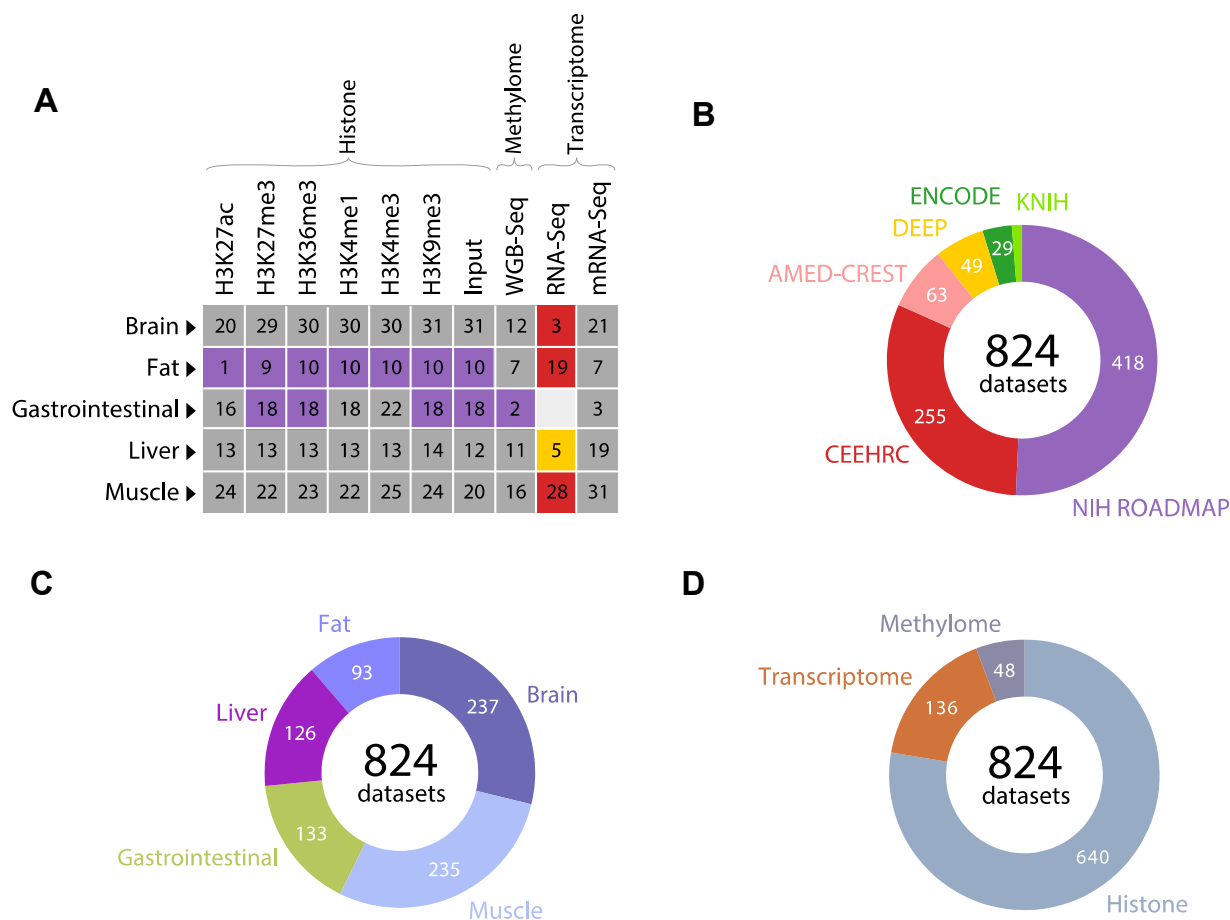


Figure 2: Metabolic disease-linked tissue epigenome maps from the International Human Epigenome Consortium (IHEC). Publicly accessible epigenome profiles from the IHEC portal (<http://epigenomes.ca/ihec>;2018-10; hg19 reference). The 824 available datasets are summarized (A) in table format, (B) per consortium, (C) per tissue type, and (D) per epigenetic mark category. Figure panels were adapted from the IHEC portal graphical content.

relationship between epigenome traits, especially at regulatory elements [16,18]. Active promoter and enhancer elements are characterized by an open chromatin profile, allowing transcription factors to access the genome and regulate transcription. By mapping open chromatin across human tissues and cells, we have been able to gain crucial insight into gene regulation as well as characterize regulatory elements in greater detail. These open chromatin regions are sensitive to cleavage by the enzyme DNase I and, consequently, marked by DNase I hypersensitivity sites (DHS) [37]. Genome-wide mapping of DHSs is performed by DNase-Seq and has provided the field of functional genomics with a reproducible tool to detect these nucleosome-free (i.e., open and active) DNA regions [38]. The application of DNase-Seq in over 125 types of human samples was a landmark study by ENCODE that showed enhancers displaying cell-type-specific patterns in contrast to promoter elements, which appeared to be largely shared across cell types and states [39]. Combining these profiles with those generated by the NIH Roadmap showed that GWAS SNPs are specifically enriched in DHS mined from human cell types linked to the disease model [17], indicating the importance of comprehensively capturing regulatory DNA in common human disease studies. More recently, the Tn5 transposase was shown to have similar abilities as DNase I in tagging and mapping open chromatin in an efficient manner, which resulted in the development of

the assay for transposase-accessible chromatin using sequencing (ATAC-Seq). Since then, ATAC-Seq has enabled investigators to assess reproducible chromatin profiles in previously unexplored cell types from lower yield samples [40]. Due to its sensitivity and ease of use over DNase-Seq, ATAC-Seq is now commonly applied in large-scale population-based studies. For instance, a recent study profiling chromatin accessibility in lymphoblastoid cell lines from 100 individuals provided insight into establishing correlation maps between open (i.e., active) chromatin regions, which may reflect three-dimensional (3D) chromatin organization in cells [41].

Post-translational modifications of histones (i.e., histone marks) represent another type of epigenetic trait studied to assign function genome-wide. Reports have shown that histone marks are deposited in a coordinated fashion (i.e., histone code) that correlates to chromatin states [42]. For instance, mono-methylation at the 4th lysine (H3K4me1) and acetylation at the 27th lysine residue of the histone H3 protein (H3K27ac) are often associated with active enhancers [43], whereas tri-methylation at the 4th lysine of the histone H3 protein (H3K4me3) correlates with active promoter regions [44]. In contrast, inactive chromatin has been noted to be depleted of these marks and, instead, demarcated by other histone modifications such as trimethylation of H3K27 (H3K27me3). In addition, silent genes have been associated with di- and trimethylation of H3K9 (i.e., H3K9me2/3)

Review

[45–48]. Using an unsupervised statistical model [15] to predict chromatin states based on these histone marks, Ernst et al. were able to show that GWAS SNPs are enriched in enhancer regions of cell types specifically linked to the associated disease or trait [49]. Focusing on the metabolic disease aspect, SNPs correlated to blood lipids and triglycerides were shown to map preferentially to distal regulatory regions defined for hepatocellular carcinoma (HepG2) cells [49]. More recently, similar integrational approaches have been applied combining both open chromatin and histone mark signatures in large meta-analyses of metabolic disease GWAS. For instance, Mahajan et al. fine-mapped type 2 diabetes loci at single-variant resolution [50] using pancreatic islet-specific chromatin profiles generated through ATAC-seq and chromatin states inferred from combinations of histone marks [51–53]. Similarly, Inoue et al. used ATAC-seq and H3K27ac ChIP-seq-derived epigenomic maps generated in leptin-responsive neurons to fine-map obesity-related GWAS SNPs [54]. Whereas these previously discussed epigenetic traits and related assays represent qualitative measures of genome function, DNA (CpG) methylation offers a quantitative measure (between 0 and 100%) at almost 30 million sites genome-wide. This feature accounts in part for the popularity of DNA methylation in epigenome-wide studies of complex traits. As such, we focus the remainder of this review on DNA methylation and complex disease studies.

4. COORDINATED PATTERNS OF METHYLOMES DEFINE REGULATORY REGIONS

As previously discussed, genome function, with its regulatory regions, can be defined by chromatin activity and histone modifications, but can also be inferred from DNA methylation patterns by taking methylation level and CpG density into account [18,39]. Regions of low to unmethylated CpGs—so-called hypomethylated footprints—are known to correspond with transcription factor occupancy and active chromatin states [55,56]. Specifically, dense regions of CpGs together with very low CpG methylation state often occur close to transcription start sites and, thus, reflect promoter regions. In contrast, regions of low CpG density together with intermediate CpG methylation overlap H3K27ac marks, linking them to enhancer regions. These enhancer-like hypomethylated footprints have further been found to show high overlaps with cell-type-specific DHS and transcription factor binding motifs [56], supporting their validity. Mechanistically, Yin et al. have recently shown that the hydrophobic nature of the methyl group at methylated CpGs is an important contributing selection factor in promoting or deterring transcription factor binding affinity [57]. To assign CpG methylation signatures to regulatory elements genome-wide, statistical models have been developed for use on next-generation single-base resolution methylation profiles. For instance, MethylSeekR uses a Hidden Markov Model to identify unmethylated (UMR) and low-methylated regions (LMR) based on average methylation status and total CpG density [56]. Studies have shown that UMRs are more likely to be shared across cell types or tissues, whereas LMRs are more prone to be tissue-specific [56,58]—reflecting the expected behaviors of promoters and enhancers, respectively. Specifically, Busche et al. investigated hypomethylated footprints within methylomes of ~30 matched subcutaneous adipose and whole-blood tissues, identifying ~60,000 regulatory regions per tissue with 30% being UMRs and the remainder LMRs [58]. UMRs were found to exhibit longer fragment lengths (~2400 bp) and 90% shared identity across tissues, whereas LMRs were found to be shorter (~750 bp) and more tissue-specific (45% shared) [58]. Importantly, identified adipose hypomethylated footprints depicted distinguishing enrichments for histone marks derived from

human adipocytes (NIH Roadmap); UMRs were enriched for H3K4me3 marks, whereas LMRs were more likely to overlap H3K4me1 marks [58]. These reported coordinated patterns of epigenetic marks indicate that inferring regulatory elements from methylation profiles alone (i.e., without the need for parallel assessments of histone modifications) is a valid strategy in the analysis of the functional epigenome in disease studies.

5. ASSESSING VARIATION OF METHYLOMES IN METABOLIC DISEASE STUDIES

Surveys of genome-wide methylome signatures profiled across various human tissues and cell types as well as in populations have demonstrated that a large fraction of the methylation landscape (up to 80%) is hypermethylated [59] and displays a relatively static pattern [59,60]. These invariable CpGs were also shown to co-localize outside of regulatory regions, further distinguishing them from the subset (~15–20%) of CpGs with dynamic regulation, which are hypomethylated, map mainly to distal regulatory elements (i.e. enhancers) and transcription factor binding sites, as well as associate with tissue-specific functions [55,59]. Schultz et al. provided further evidence of this in their genome-wide methylome survey of 18 different human tissues across four individuals, showing, for example, that hypomethylated regions of dynamic CpGs in the aorta overlap aorta-specific enhancers [60]. Similarly, a large-scale investigation of adipose methylation profiles from ~600 twins of the Multiple Tissue Human Expression Resource (MuTHER) confirmed this CpG variability pattern, with dynamic CpGs observed to be enriched within intergenic regions including those marked by H3K4me1 in human adipocytes [33]. The study also presented evidence that variable CpGs in adipose tissue are more likely to be under genetic regulation, with a subset of methylation-linked SNPs overlapping GWAS SNPs of metabolic phenotypes [33]. A similar enrichment of genetic regulation at population-variable CpGs was reported by Schultz et al., showing a ~1.6-fold enrichment of SNPs associating with dynamic regions of CpG methylation. Complementary to this, a more recent investigation of hypomethylated footprints in ~200 human visceral adipose tissue samples found that metabolic trait-linked CpGs not only map preferentially to enhancer regions, but are under strong genetic effects, with this genetic regulation being distinctively augmented at signals shared across tissue types to whole blood [35]. The aforementioned efforts strongly point to the importance of considering genomic background for functional interpretation of CpG methylation in relation to disease [61]. Specifically, full investigation of regulatory elements in disease-linked tissues will provide additional insight into complex disease etiology, prompting the implementation of targeted approaches for the assessment of methylome variation in disease association studies as discussed in the following section.

6. METHODS TO ASSESS GENOME-WIDE METHYLOME PROFILES

With the increased interest in linking epigenome profiles to disease status or trait variation in populations, various approaches suitable for large-scale assessments have been presented. Current methods for genome-wide assessment of individual CpG methylation status can be distinguished by their profiling methodology (i.e., microarray or next-generation sequencing (NGS)-based) and their genome coverage capacity (i.e., targeted vs. genome-wide). Whole-genome bisulfite sequencing (WGBS) is an NGS method that represents the gold standard in methylation assessment, wherein all CpGs (~30M CpGs in the

human genome) may be captured at single-base resolution, depending on the sequencing depth. Although the application of this method represents an advantageous and unbiased solution, the high cost of profiling the entire methylome currently limits the use of WGBS in large epigenome-wide association studies (EWAS). As stated previously, WGBS across human tissues and in populations has demonstrated that only a fraction of the methylation landscape (~20%) is variable [59], supporting a more practical solution for the assessment of the informative fraction of the methylome. In response to the high costs of WGBS, reduced representation bisulfite sequencing (RRBS) was presented as an alternative sequence-based method for methylation profiling. This enrichment-based methodology enables researchers to capture fragments from WGBS libraries that show affinity to the MspI restriction enzyme. Due to the binding motif of this enzyme, RRBS enriches for CpG-dense regions, such as promoters, CpG islands, and genic regions, at the cost of coverage at distal regulatory regions. A recent RRBS-based EWAS of 32 metabolic disease-related measurements within the population-based Metabolic Syndrome in Men (METSIM) Study revealed novel but limited number of disease-linked epigenetic regions ($N = 21$) in subcutaneous adipose tissue [62]. In line with known limitations of RRBS, a higher proportion of regions were observed to map to intragenic regions over intergenic regions. Nevertheless, in concordance with current epigenomic trends, a large fraction of the regions associated with the tested metabolic disease traits (i.e., 6 of 21) were found to overlap enhancer-like adipocyte-derived histone marks from the NIH Roadmap (i.e., H3K4me1 and H3K27ac) [62]—regions that were overall underrepresented in this study due to the preferential capture of dense CpG regions by RRBS. A more popular method that has been applied in large-scale methylation profiling studies of common disease traits is the microarray-based Infinium Human Methylation450 Beadchip (450K) array [63], which was recently replaced by the more comprehensive Infinium MethylationEPIC (EPIC) array [64]. These array methods depend on the hybridization capability of DNA followed by single-base extension and assessment at ~480,000 and ~850,000 CpG sites, respectively. Although very practical and cost-efficient to implement in the laboratory, the lack of CpG coverage and tissue-specificity provided by these “one-size-fits-all” methods can be limiting in disease studies. In fact, genomic annotation analyses have shown that these technologies are similar to RRBS, in that they are biased to CpGs located in densely populated promoter regions, which, as indicated previously, are prone to have less variable methylation levels. Conversely, distal regulatory regions containing dynamic CpGs are underrepresented on these arrays, indicating that these technologies do not permit the full assessment of the disease-linked methylome [33,35,59]. More precisely, the 450K array probes were selected to interrogate primarily elements of the RefSeq genes including CpG Islands and nearby shores and shelves, 5'UTR, 3'UTR, as well as gene bodies — together representing ~75% of the total panel [33,63]. The EPIC array builds on the 450K array by retaining ~94% of targeted sites from this method while aiming to profile additional distal regulatory elements defined by the FANTOM5 and ENCODE [64]. A 2016 report on the EPIC array method showed that 58% of probes on the chip interrogated CpGs located to FANTOM5-defined enhancers, but only 7% mapped to ENCODE distal regulatory regions. Still, these numbers represent an improvement of coverage at these types of elements over the 450K array, where only an 18% and 3% overlap were noted, respectively [64]. Congruently, in their assessments of ~300 human skeletal tissues, Taylor et al. remarked that profiling with the EPIC array limits the capacity to capture tissue-linked CpGs for use in cellular deconvolution analyses [65]. All the more, the majority of EWAS investigating

metabolic traits have relied on these user-friendly array-based technologies. These methods have been applied predominantly to cohorts of bioavailable tissues (i.e., whole-blood), providing information on a limited number of trait-linked CpGs. For instance, in their 2014 report, Dick et al. presented a novel sequential replication scheme ($N = 2500$) for an epigenome-wide study of BMI in both bioavailable and disease-linked proxy tissues (subcutaneous adipose tissues) [66]. Their results highlighted a single regulatory intronic element mapping to *HIF3A* as being linked to the disease model [66]. A larger study released in 2017 by Mendelson et al. ($N = 8000$ whole-blood samples) was able to mine a total of 83 BMI-linked CpGs enriched in DHS regions and enhancer regions marked by H3K4me1, thereby indicating the importance of profiling these regions in disease investigations [67]. These trends were confirmed by Wahl et al. in their epigenome-wide assessment of whole-blood methylation and BMI in over 10,000 individuals where, again, the trait-linked CpGs were found to be enriched in active chromatin marks [68]. Furthermore, studies focusing on correlating circulating lipids to methylation status in whole-blood tissues have demonstrated similar findings with the largest survey ($N = 4500$) by Hedman et al. identifying 33 disease-linked CpGs lacking enrichment in promoter regions [69].

To resolve some of the technical limitations posed by the aforementioned methylome assessment methods, MethylC-Capture Sequencing (MCC-Seq) was recently introduced as an equally sensitive but more economical and tissue-specific tool [34]. MCC-Seq is a NGS-based method that allows for enrichment of WGBS libraries over user-defined regions (up to 200 Mb) [34]. In this way, MCC-Seq can be used to strategically capture variable CpGs mapping to tissue-specific regulatory regions, thereby allowing users to profile informative methylomes for disease investigations while incurring comparable costs to available array-based techniques [34]. Contrary to other available capture-based methylome sequencing approaches such as Agilent SureSelect Human Methyl-Seq, MCC-Seq is able to capture both strands over selected regions, providing precise genotype calls for downstream imputation of common genotypes [34]. In this way, MCC-Seq can be applied as a dual-purpose method that permits concurrent profiling of methylation and genotypes over target regions in a tissue of interest. Although MCC-Seq can be customized for any tissue, and has been successfully applied in sperm [70] and whole-blood tissue [71], the approach was exemplified specifically for metabolic disease studies with a target panel of 4 million CpGs mapping mainly to regulatory elements specific to adipocytes and/or adipose tissue [34,35], which is discussed in more detail below.

7. FINE-MAPPING FUNCTIONAL METHYLOMES IN METABOLIC DISEASE STUDIES

Targeted investigations of the active (i.e., functional) portion of the methylome in metabolic-disease-linked tissues have provided us with additional insight into the role of regulatory regions in disease etiology. An epigenome study of 200 visceral adipose tissue samples derived from individuals with severe obesity but different degrees of metabolic complications was recently presented [35]. Here, active regulatory regions in adipose tissue and related cells were targeted by a custom panel using MCC-Seq profiling over 3 million CpGs for methylation analysis [34,35]. Intersecting these MCC-Seq-captured regulatory regions with those from available array-based methods (i.e., EPIC and 450K) found that only a small proportion of these sites are covered by these latter techniques (Figure 3). More precisely, the EPIC array captures only ~20% and ~30% of the enhancer- and promoter-mapping CpGs targeted by this adipose MCC-Seq panel, respectively

Review

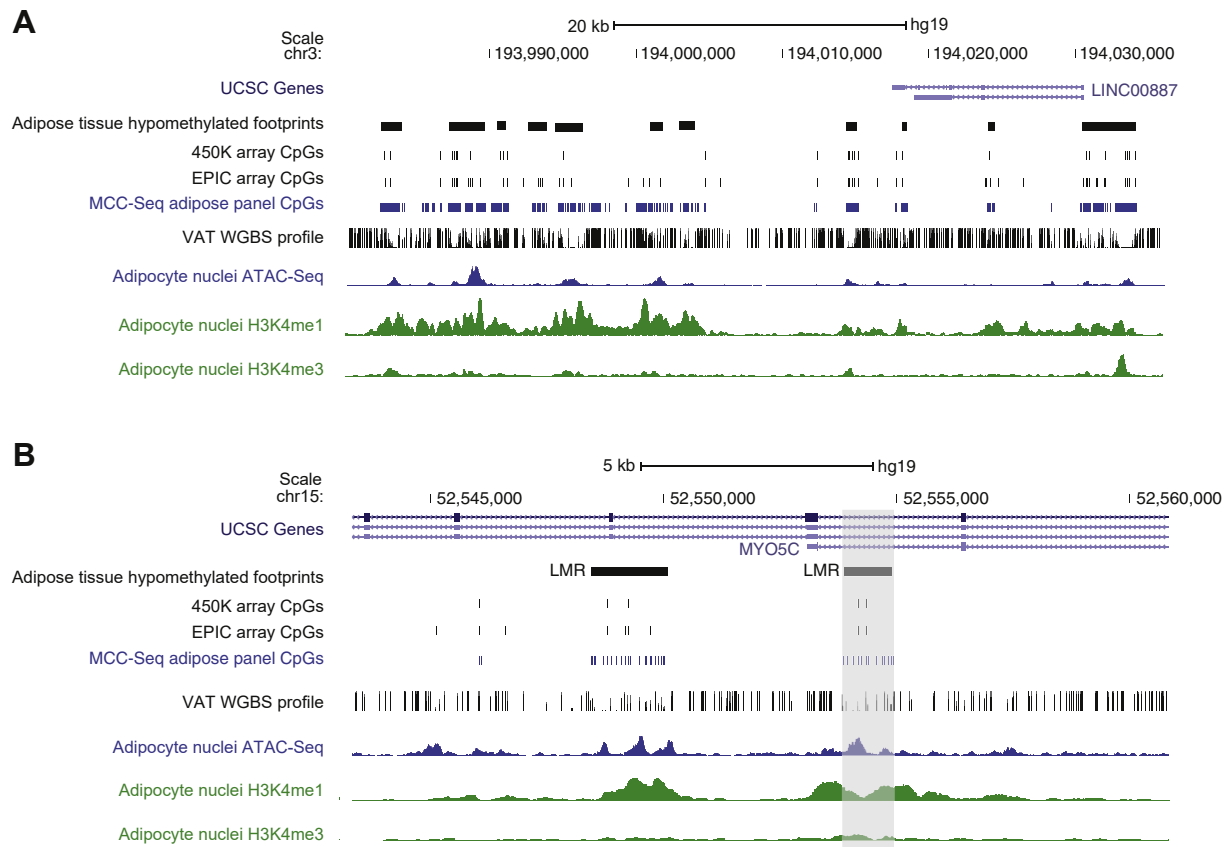


Figure 3: Comparison of DNA methylation profiling at adipose regulatory regions by available methods. CpGs captured by the 450K array, EPIC array, and MCC-Seq using the adipose tissue panel published in [34] are contrasted at different genomic regions. Overlaps with typical profiles of WGBS in visceral adipose tissue and adipocyte nuclei histone marks for H3K4me1 and H3K4me3 (NIH Roadmap; donor 92) are also depicted. (A) Lack of coverage by array-based methods can be noted at multiple adipose enhancer elements mapping to an intergenic region on chr13. (B) The fine-mapping potential by MCC-Seq at the previously identified HDL-linked 450K array probe [69] locating to an intragenic region in the *MYO5C* locus (highlighted in light gray) is exemplified.

[35]. The benefit of dense CpG methylation profiling in regulatory elements was reflected in the release of expanded catalogues of adipose regulatory regions (>500 elements per study) associating to metabolic disease traits [i.e., triglycerides, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and total cholesterol], representing higher discovery power over published studies [34,35]. As indicated previously, earlier reports of complex-trait-linked EWAS depicted disease-associated CpGs to map to enhancer-like genomic regions, but methods used to generate these findings provided only limited surveys of these types of regulatory regions (Figure 3). The application of MCC-Seq in adipose tissue permitted the investigation of the full range of distal regulatory regions as well as promoter regions mined from bioinformatics tools of whole-genome adipose DNA methylation profiles (Figure 3). This type of detailed assessment solidified observed preliminary reports, as metabolic disease trait-linked CpGs were found to be distinctively enriched not only in enhancer regions, but further in tissue-specific regulatory regions mapping to both enhancers and promoters [34,35], both of which are underrepresented in current available techniques.

The unprecedented resolution of CpG profiling by MCC-Seq also permitted the generation of distinctive positional patterns of metabolic disease trait-linked CpGs within different types of regulatory elements. Specifically, disease-associated CpGs mapping to enhancers were enriched at the mid-point of these regulatory regions, whereas those

co-localizing to promoters depicted a bimodal distribution, with TSS being less occupied [35], likely reflecting specific transcription factor (TF) occupancy at these elements. In fact, Grossman et al. used epigenome maps from 47 cell types defined by the NIH Roadmap and described similar positional trends within enhancers for different classes of TFs where those associated with cell-type-specific expression indeed mapped to the center of enhancers [72]. In line with this, Allum et al. showed that metabolic disease-associated regulatory regions were found to be TF binding motif hotspots for adiposity-linked STAT family proteins and *NFIB*. These groups of TFs were also found to depict an adipose-specific expression pattern [35]. Allum et al. further highlighted the benefit of CpG methylation profiling at single-base resolution for fine-mapping epigenetic disease loci by contrasting associations with similar adipose tissue profiling effort by the 450K array [35]. Of the replicated disease loci where at least two CpGs were covered by MCC-Seq, including one overlapping a significant metabolic disease CpG from the array-based study, ~90% of the top associating CpGs were found to locate proximally but not directly overlapping the array-based disease CpGs [35].

Investigations by Cheung et al. using MCC-Seq in a range of tissues, including visceral adipose tissues, further demonstrated that a large fraction of CpG methylation under genetic regulation is tissue-specific (>50%) and maps preferentially to enhancer elements [71]. This estimate of the genetic impact on methylation was confirmed for disease-linked CpGs (>55%) identified from visceral adipose tissue

profiles [34,35]. Importantly, complex trait-associated CpGs discovered in the disease-relevant adipose tissue that additionally replicated across tissues to the bioavailable whole-blood tissue, using matched samples, were found to exhibit even stronger genetic regulation (>90%) [35]. This observation indicates that the full depth of epigenetic impact on disease cannot be properly captured when employing cross-tissue validations from bioavailable to disease-linked tissues, as is currently being applied in many reported EWAS. Validation across well-phenotyped biologically relevant tissues of the same tissue type is needed to further explore and confirm current observed trends of complex trait-linked epigenetic variants. Allum et al. further provided functional insight for a fraction of lipid-linked GWAS SNPs, which were found to be enriched in SNP-CpG pairs overlaying adipose tissue regulatory regions that mapped with blood lipid-associated CpGs [35]. For instance, the authors highlighted HDL cholesterol-linked SNPs [73] mapping to an intragenic region of the metabolic disease *GALNT2* locus that exhibited putative regulatory effects on a nearby adipose tissue-specific enhancer element. This enhancer region was found to harbor HDL cholesterol-associated CpGs not detectable through the 450K array method [35]. These findings exemplify the advantage of using integrational approaches of association studies at single-base resolution to permit functional annotation of GWAS SNPs identified from large-scale efforts.

8. CAVEATS IN METHYLOME-BASED DISEASE ASSOCIATION STUDIES

Although epigenetic signals are known to be cell-specific, most published epigenome-wide studies to date have investigated correlations to complex traits in heterogeneous tissues. Although laboratory and statistical methods for cellular decomposition of whole blood samples are currently being applied, these are not easily transferrable to primary tissues such as adipose tissue, where many unknowns remain in terms of cellular makeup. To account for tissue heterogeneity in adipose EWAS, investigators have predominantly integrated BMI measures as a surrogate to indicate levels of adipose tissue heterogeneity [34,35,74]. Although BMI does not necessarily mirror body fat distribution, Laforest et al. have shown that across multiple techniques tested for adipose cell size, a general pattern of association between this measure and BMI is noted — indicating that BMI can be used as a proxy for cellular heterogeneity [74].

Nevertheless, more precise estimates that better reflect an individual's adipose tissue composition are warranted to account for the impact of tissue heterogeneity in EWAS. For instance, we know that visceral adipose tissues of obese individuals undergo important remodeling and infiltration by immune cell types such as macrophages, but estimates are hard to evaluate in large-scale studies. Efforts from large consortia are providing more and more reference epigenome maps to be used for tissue dissection purposes. Targeted purified primary cell-based studies, such as the one from Bradford et al. that contrasts epigenome profiles of adipocytes from different fat depots [32], further provide insight to the community to help pinpoint observed signals and decipher how epigenetic changes translate mechanistically. Similarly, studies using single-cell technologies are crucial to elucidate cellular components of disease-linked primary tissues. A recent report by Vijay et al. used single-cell RNA-Seq to profile the stromal fraction of 12 visceral and 13 subcutaneous adipose tissue samples, enabling the discovery of crucial differences in cellular composition between these fat depots [75]. Together, these efforts will allow for the implementation of reference-based or supervised deconvolution methods in complex tissue-based EWAS similar to what is already available for studies of whole blood

[76]. Similarly, efforts to improve reference-free/unsupervised deconvolution methods will also be beneficial to help correct currently used statistical models in disease dissection. Briefly, these methods use an approach that does not require any additional input of reference methylomes to determine cofactors from methylation levels that correspond to cellular components of tissues among others. These cofactors can then be added to the statistical model to correct for tissue heterogeneity and other components. As previously stated here, most methods developed for this purpose have been benchmarked in whole blood samples, whereas their application in complex primary tissues remains to be validated. For instance, the publicly available Reference-Free Adjustment for Cell-Type composition (ReFACTor) software was released in 2016 and uses principal component analyses on a subset of differentially methylated regions to provide approximates of variables to be used in EWAS models [77]. While the expectation is that top variables pulled out by the software will correspond to cellular components of the tissue, testing the technique within cancerous breast tissues showed that variables corresponding to other components such as disease status were instead highlighted [78]. Another technique developed in 2016 by Houseman et al. called RefFreeCellMix is based on naive matrix factorization (NMF) and was benchmarked on methylation levels from multiple tissue cohorts profiled on Illumina arrays [79]. More recently, MeDeCom, which uses NMF as with RefFreeCellMix, was released; however, it also applies constraints based on DNA methylation behavior in pure cell populations to improve latent methylation component prioritization [80]. In their proof-of-concept paper, the authors showed that MeDeCom outperformed RefFreeCellMix when identifying cellular components in more complex samples [80]. Extrapolation of these algorithms for use on single-base resolution methylomes as opposed to array-based data will be useful in future epigenomic studies.

As previously stated here, epigenetic traits are known to be modulated by both environmental and genetics effects and, thus, can be seen as reflecting the interplay between these drivers of complex disease. However, an important question remains in terms of dissecting the exact contribution of these two modules to epigenetic mark variation. For example, in instances of integrative data gathering (i.e., genotypes, epigenetic traits, and phenotypes), the user can dissect genetic and potential environmental contributions to the identified epigenetic signatures by conditioning association models on the top-linked SNPs [35]. This method is particularly useful when investigating epigenetic trait data in population-wide studies, and more powerful when gathering information at single-base resolution. However, to fully account for genetic confounding factors in studies seeking to identify epigenetic modulations associated with disease, a discordant twin model of monozygotic twins is ideal [33,81–83].

Finally, a consequence of generating epigenome maps in cross-sectional cohorts is that it does not permit for causality inference of the epigenetic change on complex traits (Figure 1). While longitudinal study models and functional validation would be ideal to investigate this type of relationship, statistical models have been developed to infer the direction of the causal link between associated epigenetic and complex traits. The most commonly applied of these models is Mendelian randomization (MR). Briefly, MR uses an instrumental variable (IV) in the form of SNPs linked only to the exposure under study, given that they are not prone to reverse causation, to determine the direction of effect between the exposure and the outcome. Relating to epigenetic traits such as DNA methylation, causal effects may only be explored at CpGs exhibiting genetic linkage. Applying the statistical method of MR, Dekkers et al. substantiated findings that support an effect of circulating lipid status on correlated methylation levels [84]. Similar findings were reported by Mendelson et al. for the influence of BMI on

Review

methylation levels at a proportion of identified BMI-linked CpGs [67]. Whereas in their large-scale methylome study of 139 complex traits, Richardson and colleagues were unable to resolve causality with MR due to horizontal pleiotropy, MR findings were used to prioritize potential mediation effects by GWAS SNPs on gene expression through DNA methylation for over 300 CpG-SNP pairs [85]. However, all these findings were gathered from whole-blood DNA methylation and were not corroborated in disease-relevant tissues, making translation into biological knowledge challenging. While Wahl et al. have made strides in this aspect by presenting similar results for the impact of BMI on CpG methylation in subcutaneous adipose tissue [68], further investigation in more tightly disease-linked tissues is needed to confirm this trend. Furthermore, given that these studies used the 450K array, the direction of causality of methylation in metabolic disease studies at regulatory regions lying outside of promoters (i.e., distal enhancer elements) has not yet been evaluated. These assessments will be crucial in future complex disease etiology studies.

9. CONCLUSIONS AND FUTURE DIRECTIONS

Fast-paced progress has been made in the field of functional epigenomics, both in terms of technological advancement to study epigenomes, as well as our ability to apply these technologies in large cohort studies. Functional epigenome profiles exhibit tissue and cell-specificity, yet most epigenome-wide studies of metabolic traits have been conducted in whole-blood tissues due to easy access, as opposed to tissues that are more targeted to the disease model such as visceral adipose or liver tissues. Whole-genome methylation profiling across a large number of human tissues has confirmed that enhancer regions—known to play a role in gene expression modulation—are predominantly tissue- and cell-specific [58]. Comprehensive studies correlating methylation status to metabolic traits using array-based technologies biased toward promoter regions have provided significant but restricted insight into disease etiology, due to the static nature of these elements. The advent of targeted single-base resolution technologies, such as MCC-Seq, coupled with the availability of well-phenotyped clinically relevant cohorts have provided increased discovery power for the detection of biologically relevant trait-linked epigenetic variants.

An important point of future exploration to assess the full contribution of epigenetic variants in metabolic disease outcome will be to progress toward assessing the ability of epigenetic marks to predict future disease risk. In fact, a recent longitudinal EWAS study in more than 10,000 individuals reported by Agha et al. showed the predictive nature of methylation status on myocardial infarction onset at more than 50 CpGs detectable on the 450K array—indicating that methylation status can be used as an indicator of future complex disease development [86].

Moreover, population-based epigenome-wide studies have been mostly centered on linking variation to disease models. However, the issue remains that many metabolic disease-related traits, such as fat distribution, depict sexual dimorphism [87–89], which remain uncharted in terms of epigenome maps. Large-scale well-phenotyped cohorts within disease-targeted tissues extracted from both males and females are needed to study these effects in more detail and permit us to better treat metabolic disease in the dimorphic human population. This will be an interesting area of future exploration.

Additional high-resolution integrational approaches to link identified disease-associated epigenome changes to target genes are needed. Here, the advent of chromatin conformation and editing technologies in disease-linked cell types are promising [90]. For instance, efforts from

the 4D Nucleome project are ongoing to generate reliable genomic interaction profiles in obesity-related cell types such as differentiated adipocytes. Finally, profiling novel histone modifications, such as propionylation and butyrylation, in metabolic disease-linked tissue models will allow additional insight into epigenome perturbation and disease susceptibility [91].

FUNDING

E.G. holds the Roberta D. Harding & William F. Bradley, Jr. Endowed Chair in Genomic Research. F.A. held a studentship from the Fonds de Recherche en Santé du Québec (FRSQ) during part of this study and is currently supported by a team grant from the Canadian Institutes of Health Research (CIHR) as part of the 4D Nucleome Transformative Collaborative Project.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

REFERENCES

- [1] Kilpinen, H., Dermitzakis, E.T., 2012. Genetic and epigenetic contribution to complex traits. *Human Molecular Genetics* 21(R1):R24–R28.
- [2] Barres, R., Zierath, J.R., 2011. DNA methylation in metabolic disorders. *The American Journal of Clinical Nutrition* 93(4):897S–900S.
- [3] Gluckman, P.D., Hanson, M.A., Buklijas, T., Low, F.M., Beedle, A.S., 2009. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nature Reviews Endocrinology* 5(7):401–408.
- [4] Elder, S.J., Lichtenstein, A.H., Pittas, A.G., Roberts, S.B., Fuss, P.J., Greenberg, A.S., et al., 2009. Genetic and environmental influences on factors associated with cardiovascular disease and the metabolic syndrome. *Journal of Lipid Research* 50(9):1917–1926.
- [5] McCarthy, M.I., 2010. Genomics, type 2 diabetes, and obesity. *New England Journal of Medicine* 363(24):2339–2350.
- [6] González-Muniesa, P., Martínez-González, M.-A., Hu, F.B., Després, J.-P., Matsuzawa, Y., Loos, R.J.F., et al., 2017. Obesity. *Nature Reviews Disease Primers* 3:17034.
- [7] Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., et al., 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518(7538):197.
- [8] Consortium, t.G., Kemper, K.E., Zheng, Z., Yengo, L., Sidorenko, J., Yang, J., et al., 2018. Meta-analysis of genome-wide association studies for height and body mass index in ~700,000 individuals of European ancestry. *Human Molecular Genetics* 27(20):3641–3649.
- [9] Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., et al., 2015. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518(7538):187.
- [10] Stunkard, A.J., Harris, J.R., Pedersen, N.L., McClearn, G.E., 1990. The body-mass index of twins who have been reared apart. *New England Journal of Medicine* 322(21):1483–1487.
- [11] Allison, D.B., Kaprio, J., Korkeila, M., Koskenvuo, M., Neale, M.C., Hayakawa, K., 1996. The heritability of body mass index among an international sample of monozygotic twins reared apart. *International Journal of Obesity Related Metabolic Disorders: Journal of the International Association for the Study of Obesity* 20(6):501–506.
- [12] Andersen, M.K., Sandholt, C.H., 2015. Recent progress in the understanding of obesity: contributions of genome-wide association studies. *Current Obesity Reports* 4(4):401–410.

- [13] Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., et al., 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics* 47(10):1114–1120.
- [14] Zuk, O., Hechter, E., Sunyaev, S.R., Lander, E.S., 2012. The mystery of missing heritability: genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America* 109(4):1193–1198.
- [15] Ernst, J., Kellis, M., 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology* 28(8): 817–825.
- [16] Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., et al., 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317.
- [17] Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., et al., 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099):1190–1195.
- [18] Consortium, E.P., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
- [19] Romanoski, C.E., Glass, C.K., Stunnenberg, H.G., Wilson, L., Almouzni, G., 2015. Epigenomics: Roadmap for regulation. *Nature* 518(7539):314.
- [20] Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., et al., 2011. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genetics* 7(2):e1002003.
- [21] Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., et al., 2012. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* 44(10):1084–1089.
- [22] Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., et al., 2016. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* 167(5):1398–1414 e1324.
- [23] GTEx, C., 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235):648–660.
- [24] Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., et al., 2018. Using an atlas of gene regulation across 44 human tissues to inform complex disease-and trait-associated variation. *Nature Genetics* 50(7): 956.
- [25] GTEx, C., 2017. Genetic effects on gene expression across human tissues. *Nature* 550(7675):204.
- [26] Collins, F.S., Morgan, M., Patrinos, A., 2003. The Human Genome Project: lessons from large-scale biology. *Science* 300(5617):286–290.
- [27] Group, I.S.M.W., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409(6822):928.
- [28] Consortium, E.P., 2004. The ENCODE (ENCyclopedia of DNA elements) project. *Science* 306(5696):636–640.
- [29] Consortium, E.P., 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biology* 9(4):e1001046.
- [30] Mapping the epigenome. *Nature Methods* 12, 2015:161.
- [31] Stunnenberg, H.G., Abrignani, S., Adams, D., de Almeida, M., Altucci, L., Amin, V., et al., 2016. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell* 167(5):1145–1149.
- [32] Bradford, S.T., Nair, S.S., Statham, A.L., van Dijk, S.J., Peters, T.J., Anwar, F., et al., 2019. Methylome and transcriptome maps of human visceral and subcutaneous adipocytes reveal key epigenetic differences at developmental genes. *Scientific Reports* 9(1):9511.
- [33] Grundberg, E., Meduri, E., Sandling, J.K., Hedman, Å.K., Keildson, S., Buil, A., et al., 2013. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *The American Journal of Human Genetics* 93(5):876–890.
- [34] Allum, F., Shao, X., Guénard, F., Simon, M.-M., Busche, S., Caron, M., et al., 2015. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nature Communications* 6.
- [35] Allum, F., Hedman, Å.K., Shao, X., Cheung, W.A., Vijay, J., Guénard, F., et al., 2019. Dissecting features of epigenetic variants underlying cardiometabolic risk using full-resolution epigenome profiling in regulatory elements. *Nature Communications* 10(1):1209.
- [36] Innocenti, F., Cooper, G.M., Stanaway, I.B., Gamazon, E.R., Smith, J.D., Mirkov, S., et al., 2011. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genetics* 7(5):e1002078.
- [37] Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., et al., 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132(2):311–322.
- [38] Gross, D.S., Garrard, W.T., 1988. Nuclease hypersensitive sites in chromatin. *Annual Review of Biochemistry* 57(1):159–197.
- [39] Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., et al., 2012. The accessible chromatin landscape of the human genome. *Nature* 489(7414):75.
- [40] Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* 10(12):1213.
- [41] Kumasaka, N., Knights, A.J., Gaffney, D.J., 2019. High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nature Genetics* 51(1):128.
- [42] Strahl, B.D., Allis, C.D., 2000. The language of covalent histone modifications. *Nature* 403(6765):41.
- [43] Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* 39(3):311–318.
- [44] Liang, G., Lin, J.C., Wei, V., Yoo, C., Cheng, J.C., Nguyen, C.T., et al., 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proceedings of the National Academy of Sciences of the United States of America* 101(19):7357–7362.
- [45] Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., et al., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410(6824):120.
- [46] Beisel, C., Paro, R., 2011. Silencing chromatin: comparing modes and mechanisms. *Nature Reviews Genetics* 12(2):123.
- [47] Becker, J.S., Nicetto, D., Zaret, K.S., 2016. H3K9me3-dependent heterochromatin: barrier to cell fate changes. *Trends in Genetics* 32(1):29–41.
- [48] Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., et al., 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development* 18(11):1251–1262.
- [49] Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43–49.
- [50] Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., et al., 2018. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics* 50(11):1505.
- [51] Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., et al., 2017. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* 114(9):2301–2306.
- [52] Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., et al., 2014. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nature Genetics* 46(2): 136.

Review

- [53] Thurner, M., Van De Bunt, M., Torres, J.M., Mahajan, A., Nylander, V., Bennett, A.J., et al., 2018. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife* 7: e31977.
- [54] Inoue, F., Eckalbar, W.L., Wang, Y., Murphy, K.K., Matharu, N., Vaisse, C., et al., 2019. Genomic and epigenomic mapping of leptin-responsive neuronal populations involved in body weight regulation. *Nature Metabolism* 1(4):475.
- [55] Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., et al., 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480(7378):490–495.
- [56] Burger, L., Gaidatzis, D., Schübeler, D., Stadler, M.B., 2013. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Research* 41(16):e155.
- [57] Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., et al., 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356(6337):eaaj2239.
- [58] Busche, S., Shao, X., Caron, M., Kwan, T., Allum, F., Cheung, W.A., et al., 2015. Population whole-genome bisulfite sequencing across two tissues highlights the environment as the principal source of human methylome variation. *Genome Biology* 16(1):290.
- [59] Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., et al., 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500(7463):477–481.
- [60] Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., et al., 2015. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523(7559):212.
- [61] Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* 13(7):484–492.
- [62] Orozco, L.D., Farrell, C., Hale, C., Rubbi, L., Rinaldi, A., Civelek, M., et al., 2018. Epigenome-wide association in adipose tissue from the METSIM cohort. *Human Molecular Genetics* 27(10):1830–1846.
- [63] Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., et al., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6(6):692–702.
- [64] Pidsley, R., Zotenko, E., Peters, T.J., Lawrence, M.G., Risbridger, G.P., Molloy, P., et al., 2016. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology* 17(1):208.
- [65] Taylor, D.L., Jackson, A.U., Narisu, N., Hemani, G., Erdos, M.R., Chines, P.S., et al., 2019. Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proceedings of the National Academy of Sciences of the United States of America* 116(22):10883–10888.
- [66] Dick, K.J., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aïssi, D., Wahl, S., et al., 2014. DNA methylation and body-mass index: a genome-wide analysis. *The Lancet*.
- [67] Mendelson, M.M., Marioni, R.E., Joehanes, R., Liu, C., Hedman, Å.K., Aslibekyan, S., et al., 2017. Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a Mendelian randomization approach. *PLoS Medicine* 14(1):e1002215.
- [68] Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., et al., 2017. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 541(7635):81.
- [69] Hedman, Å.K., Mendelson, M.M., Marioni, R.E., Gustafsson, S., Joehanes, R., Irvin, M.R., et al., 2017. Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies. *Circulation: Cardiovascular Genetics* 10(1):e001487.
- [70] Chan, D., Shao, X., Dumargne, M.-C., Aarabi, M., Simon, M.-M., Kwan, T., et al., 2019. Customized MethylC-capture sequencing to evaluate variation in the human sperm DNA methylome representative of altered folate metabolism. *Environmental Health Perspectives* 127(8):087002.
- [71] Cheung, W.A., Shao, X., Morin, A., Siroux, V., Kwan, T., Ge, B., et al., 2017. Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biology* 18(1):50.
- [72] Grossman, S.R., Engreitz, J., Ray, J.P., Nguyen, T.H., Hacohen, N., Lander, E.S., 2018. Positional specificity of different transcription factor classes within enhancers. *Proceedings of the National Academy of Sciences of the United States of America* 115(30):E7222–E7230.
- [73] Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., et al., 2013. Discovery and refinement of loci associated with lipid levels. *Nature Genetics* 45(11):1274.
- [74] Laforest, S., Labrecque, J., Michaud, A., Cianflone, K., Tchernof, A., 2015. Adipocyte size as a determinant of metabolic disease and adipose tissue dysfunction. *Critical Reviews in Clinical Laboratory Sciences* 52(6):301–313.
- [75] Vijay, J., Gauthier, M.-F., Biswell, R.L., Louiselle, D.A., Johnston, J.J., Cheung, W.A., et al., 2019. Single-cell analysis of human adipose tissue identifies depot- and disease-specific cell types. *Nature Metabolism*. In Press.
- [76] Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., et al., 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13(1):86.
- [77] Rahmani, E., Zaitlen, N., Baran, Y., Eng, C., Hu, D., Galanter, J., et al., 2016. Sparse PCA corrects for cell type heterogeneity in epigenome-wide association studies. *Nature Methods* 13(5):443.
- [78] Zheng, S.C., Beck, S., Jaffe, A.E., Koestler, D.C., Hansen, K.D., Houseman, A.E., et al., 2017. Correcting for cell-type heterogeneity in epigenome-wide association studies: revisiting previous analyses. *Nature Methods* 14(3):216.
- [79] Houseman, E.A., Kile, M.L., Christiani, D.C., Ince, T.A., Kelsey, K.T., Marsit, C.J., 2016. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics* 17(1):259.
- [80] Lutsik, P., Slawski, M., Gasparoni, G., Vedenev, N., Hein, M., Walter, J., 2017. MeDeCom: discovery and quantification of latent components of heterogeneous methylomes. *Genome Biology* 18(1):55.
- [81] Livshits, G., Gao, F., Malkin, I., Needhamsen, M., Xia, Y., Yuan, W., et al., 2016. Contribution of heritability and epigenetic factors to skeletal muscle mass variation in United Kingdom twins. *Journal of Clinical Endocrinology & Metabolism* 101(6):2450–2459.
- [82] Pietiläinen, K., Ismail, K., Järvinen, E., Heinonen, S., Tummers, M., Bollepalli, S., et al., 2016. DNA methylation and gene expression patterns in adipose tissue differ significantly within young adult monozygotic BMI-discordant twin pairs. *International Journal of Obesity* 40(4):654.
- [83] Ollikainen, M., Ismail, K., Gervin, K., Kyllönen, A., Hakkarainen, A., Lundbom, J., et al., 2015. Genome-wide blood DNA methylation alterations at regulatory elements and heterochromatin regions in monozygotic twins discordant for obesity and liver fat. *Clinical Epigenetics* 7(1):39.
- [84] Dekkers, K.F., van Iterson, M., Slieker, R.C., Moed, M.H., Bonder, M.J., Van Galen, M., et al., 2016. Blood lipids influence DNA methylation in circulating cells. *Genome Biology* 17(1):138.
- [85] Richardson, T.G., Haycock, P.C., Zheng, J., Timpson, N.J., Gaunt, T.R., Davey Smith, G., et al., 2018. Systematic Mendelian randomization framework elucidates hundreds of CpG sites which may mediate the influence of genetic variants on disease. *Human Molecular Genetics* 27(18):3293–3304.
- [86] Agha, G., Mendelson, M.M., Ward-Caviness, C.K., Joehanes, R., Huan, T., Gondalia, R., et al., 2019. Blood leukocyte DNA methylation predicts risk of future myocardial infarction and coronary heart disease: a longitudinal study of 11 461 participants from population-based cohorts. *Circulation* 140(8):645–657.

- [87] Poulsen, P., Vaag, A., Kyvik, K., Beck-Nielsen, H., 2001. Genetic versus environmental aetiology of the metabolic syndrome among male and female twins. *Diabetologia* 44(5):537–543.
- [88] Varlamov, O., Bethea, C.L., Roberts Jr., C.T., 2015. Sex-specific differences in lipid and glucose metabolism. *Frontiers in Endocrinology* 5:241.
- [89] Lovejoy, J., Sainsbury, A., Group, S.C.W., 2009. Sex differences in obesity and the regulation of energy homeostasis. *Obesity Reviews* 10(2):154–167.
- [90] Pan, D.Z., Garske, K.M., Alvarez, M., Bhagat, Y.V., Boocock, J., Nikkola, E., et al., 2018. Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nature Communications* 9(1):1512.
- [91] Kebede, A.F., Nieborak, A., Shahidian, L.Z., Le Gras, S., Richter, F., Gomez, D.A., et al., 2017. Histone propionylation is a mark of active chromatin. *Nature Structural Molecular Biology* 24(12):1048.