

The importance of detailed epigenomic profiling of different cell types within organs

The human body consists of hundreds of kinds of cells specified from a single genome overlaid with cell type-specific epigenetic information. Comprehensively profiling the body's distinct epigenetic landscapes will allow researchers to verify cell types used in regenerative medicine and to determine the epigenetic effects of disease, environmental exposures and genetic variation. Key marks/factors that should be investigated include regions of nucleosome-free DNA accessible to regulatory factors, histone marks defining active enhancers and promoters, DNA methylation levels, regulatory RNAs, and factors controlling the three-dimensional conformation of the genome. Here we use the lung to illustrate the importance of investigating an organ's purified cell epigenomes, and outline the challenges and promise of realizing a comprehensive catalog of primary cell epigenomes.

First draft submitted: 10 January 2016; Accepted for publication: 1 March 2016; Published online: 15 June 2016

Keywords: cell-specific epigenomes • ChIP-seq • DNA methylation • environmental epigenetics • epigenomics • histone modification • lung • organ

A single fertilized human egg, providing a single diploid genome, gives rise to a vast array of cells, recently estimated to be about 37 trillion [1]. These cells are thought to represent over 210 histologically distinct cell types, most organized in complex higher order structures that form highly specialized organs [2]. A key question in human biology is how these distinct cell types are derived and maintained during development and adulthood. Because the genomic information in all cells is identical (excepting specific genetic rearrangements such as immunoglobulin gene recombination in B cells), the distinct phenotypic properties of different cell types in an organ must be acquired by restricting accessibility to the genetic code in a cell type-specific and dynamic manner. This is accomplished by the epigenome, defined as the information overlaid on the genome that blocks access to certain regions and opens others to gene expression. Epig-

enomic information encompasses chemical modifications of DNA such as DNA methylation and hydroxymethylation, proteins bound to the DNA such as transcription factors that populate promoters and enhancers, modifications of the histones that make up the nucleosomes, nucleosome remodeling complexes, proteins that influence the three-dimensional organization of the genome such as lamins and CTCF, and numerous noncoding regulatory RNAs (Figure 1). Together, these epigenetic factors constitute the epigenome and specify and maintain the different cell types in the body. The analysis of the epigenome can provide insight into disease risk and onset when affected cell types are compared with their healthy (reference) counterparts. While the cost of sequencing may necessarily limit which features can be profiled for many different cell types, the epigenetic marks in Table 1 are widely accepted as defining regulatory regions of DNA. In

Theresa Ryan Stueve^{1,2,3},
Crystal N Marconett^{1,2},
Beiyun Zhou^{2,4}, Zea Borok^{2,3,4}
& Ite A Laird-Offringa^{*1,2,3}

¹Department of Surgery, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

²Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

³Department of Biochemistry & Molecular Biology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

⁴Division of Pulmonary & Critical Care Medicine, Department of Medicine, Will Rogers Institute Pulmonary Research Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

*Author for correspondence:

Tel.: +1 323 865 0655

Fax.: +1 323 865 0158

ilaird@usc.edu

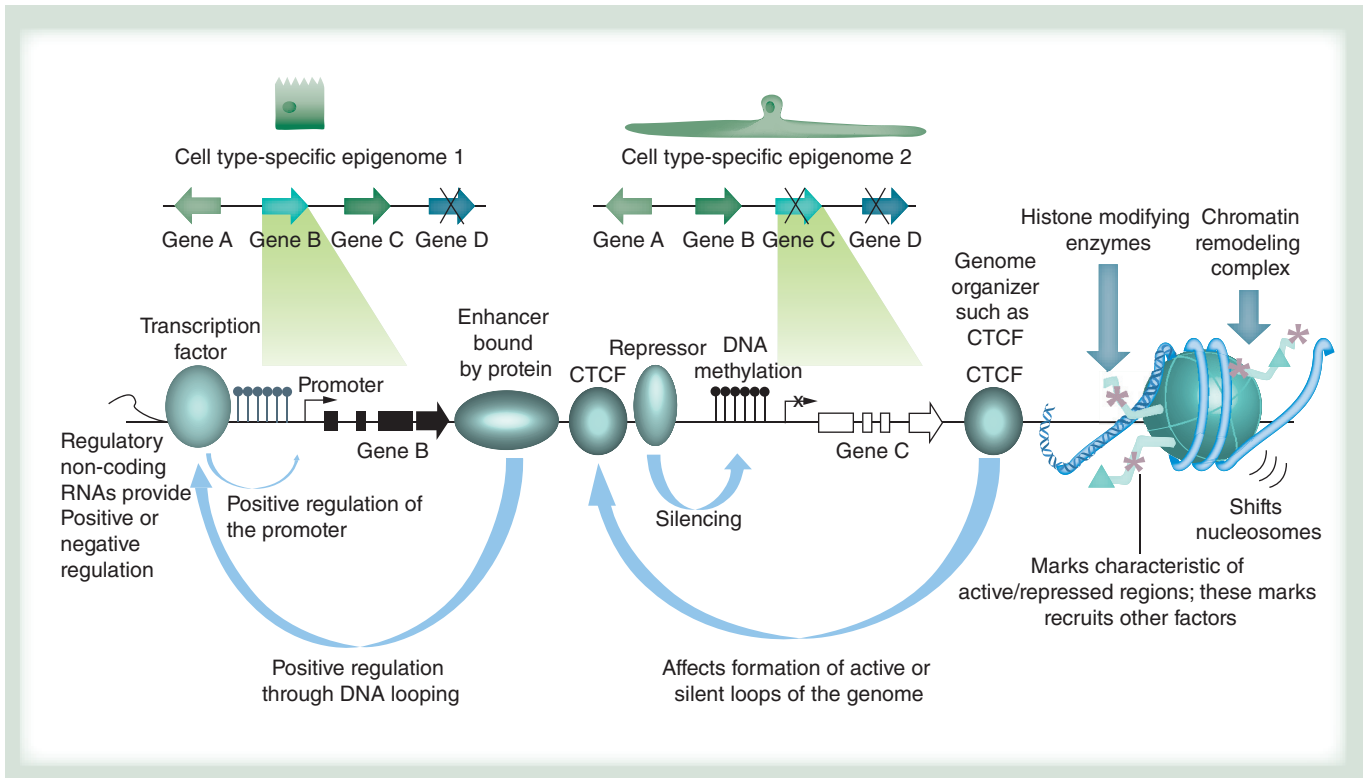


Figure 1. Mechanisms of epigenetic control. The epigenome allows cell types with identical genomes (schematically depicted at the top) to exhibit differential gene expression patterns through a variety of mechanisms. They include, from left to right: regulatory RNAs of different types such as long noncoding RNAs, miRNAs and others; recruitment of transcription factors that can activate gene expression; enhancer-binding proteins that can activate gene expression by looping to promoters in *cis* or in *trans* (when different chromosomes interact); proteins like CTCF that mediate looping interactions in concert with other proteins such as cohesin and thereby support the three-dimensional organization of the genome; the binding of repressor proteins that can silence genes; DNA methylation that is usually absent in active regulatory regions such as promoters and enhancers (hydroxymethylation, thought to occur during DNA methylation removal, and other modifications can also occur); a variety of histone modifying enzymes that lead to protein interactions associated with different chromatin states; nucleosome remodeling complexes that can lead to the removal or deposition of nucleosomes in active and repressed genomic regions, respectively.

the following sections we use the lung as an example to highlight the importance of understanding the epigenetic basis for cell phenotype specification, and describe specific instances in which cell type-specific epigenomic information might be leveraged: in tissue regeneration research, in determining disease states compared with healthy progenitor cells, in assessing the consequences of environmental exposures and in identifying functional noncoding variants associated with disease risk in genome-wide association studies (GWAS).

Examination of purified cell types is crucial for tissue regeneration research

The availability of human embryonic stem cells (ESCs) and induced pluripotent stem cells has opened avenues for *in vitro* differentiation of different tissues and cell types, so that repair and regeneration of all tissues is becoming a reality. However, derived cells and tissues must properly recapitulate the natural state and pheno-

type, or they may put patients at risk for tissue degeneration, malfunction, or possibly even cancer [12]. Analysis of the epigenomes of numerous established human ESC lines has shown abundant epigenetic abnormalities, indicating that culture conditions might predispose cells to neoplastic transformation [13]. Whether the proper phenotype of progenitors or differentiated cells derived from stem cells has been achieved can be assessed by examining whether the epigenome (and thereby the transcriptome) of the derived cells mirrors those of the desired target cell. However, this requires that the epigenomes of stem cells and target cells be well characterized. The Roadmap Epigenome project and the ENCODE project have made progress in this effort [6,13]. Among the 111 analyzed cell types in the Roadmap Epigenome project are induced pluripotent cells, human embryonic stem cells and several derived progenitor cells. However, the characterization of individual cell types that make up complex organs is in its very early stages. Due to the difficulties in purifying

Table 1. Recommended epigenetic factors to be examined.

Epigenetic factor	Technique	Rationale	Feasible for low cell number approaches?
Noncoding RNA	Purification and whole transcriptome RNA-seq adapted to different types of RNAs including miRNAs and long noncoding RNAs	Noncoding RNAs, including miRNAs and lncRNAs, are strongly implicated in gene regulation and are deregulated in disease processes. RNA-seq will simultaneously provide information on regulatory RNAs, gene expression, and splicing isoforms	Single cell RNA-seq is viable but may be very challenging for lncRNAs, many of which are lowly abundant. Single cell RNA-seq has been performed for lung cells [3] and other tissues
Open chromatin	Assay for transposase-accessible chromatin with high-throughput sequencing	This method requires fewer cells than other methods to map open chromatin, like measurement of DNase hypersensitive sites or formaldehyde-assisted identification of regulatory elements	Recently a single-cell method for ATAC-seq was published [4]
DNA methylation	Whole-genome bisulfite sequencing or nanopore-based sequencing	Genome-wide DNA methylation determination provides an unbiased assessment of the location of DNA methylation*	Single cell bisulfite sequencing has been published [5], as well as solid state nanopore-based sequencing using the methyl-binding protein MBD1 [5]
H3K4me1	ChIP-seq	A histone mark associated with active and poised enhancer regions [6]	Low-cell-number approaches include iChIP-seq [7] or recovery via protection (RP)-ChIP-seq and favored amplification RP-ChIP-seq (FARP-ChIP-seq) [8]
H3K27Ac	ChIP-seq	A histone mark associated with active enhancers and promoters [6]	As above
H3K27me3	ChIP-seq	A histone mark indicative of Polycomb repression and gene inactivity [9]	As above
CTCF	ChIP-seq	CTCF is a chromatin-organizing protein involved in looping interactions between enhancers and promoters [10], among others	Due to the lower abundance of DNA-binding factor marks, low cell assays may not yet be within reach

*Whole-genome bisulfite sequencing does not distinguish between methyl and hydroxyl-methyl C, an oxidized version of methylcytosine that is a likely intermediate in demethylation and that is common in the nervous system [11].
RNA-seq: RNA sequencing.

large numbers of individual cell types, initial efforts to elucidate organ-specific epigenomes have focused largely on the characterization of tissues samples consisting of mixed cell populations. While this provides a good starting point, the resulting data reflects an average epigenome that may be based on vastly differing epigenomes of distinct cell populations [14–16]. A limited examination of purified cell populations has taken place, including numerous immortalized or cancerous human cell lines and several fully differentiated purified cell types, such as skin keratinocytes, skin fibroblasts and melanocytes [17–19]. Although collectively these efforts have provided key insights into tissue-specific epigenomic differences, it remains largely unknown how the individual cell types within each organ contribute to the average epigenomes obtained for these tissues.

Identifying disease-associated epigenetic aberrations will require knowledge of progenitor cells & establishment of reference epigenomes

Numerous human diseases have been associated with epigenomic abnormalities, including diabetes [20,21], sepsis [15], mental illnesses [16,22–25] and all types of cancer [26]. While promising, challenges to future similar studies are at least threefold:

- Difficulties related to obtaining pure populations of diseased cells from patient tissues;
- Debate around true cells of origin for many diseases; and
- The dearth of comprehensive reference or ‘baseline’ epigenomes even in scenarios in which the progenitor cells of disease are undisputed.

Studies of cancer illustrate all three examples. First, the simple objective of obtaining pure populations of ‘tumor’ and ‘normal’ cells is complicated by the fact that the tumor mass is comprised of a heterogeneous cell mixture comprising connective tissue, blood vessels and infiltrating lymphocytes. In addition, tumor samples are often compared with adjacent noncancerous tissue in the same patient, which can also be comprised of numerous cell types. Mathematical models for dealing with tissue heterogeneity using DNA methylation signatures have been developed and can address this problem to a certain extent [14,18,19]. Heterogeneity can also be minimized by expert pathologists and accurate microdissection of clinical samples, however, this may come at the expense of sufficient starting material for downstream sequencing. Despite these

efforts, it remains a challenge to obtain purified cell types, particularly where large epidemiological studies are concerned.

Secondly, the cell type of origin for many diseases may not be known. Moreover, even if the cell type of origin is known, such cells may be rare in samples of noncancerous tissue [27,28]. The latter would be true if the cell type of origin is a stem cell present in the tissue of interest, or if it represents a very small fraction of cells in normal tissue. A prime example is small-cell lung cancer (SCLC), a very aggressive cancer found predominantly in smokers. SCLC arises from rare pulmonary neuroendocrine cells which lie in small clusters along the airways, as discussed below [29]. Even though epigenomic profiles of SCLC have been described [30,31], it is unclear which aspects are abnormal because the epigenomes of the healthy reference cell populations have not been obtained.

Measuring the effects of environmental exposures requires knowledge of the baseline epigenomes of unexposed cell types

Studies of mono- and dizygotic twins have been particularly informative in illustrating the profound effects of epigenomic variation on human health and the role of the environment in inducing this variation [25,32,33]. In Alzheimer’s disease (AD), early life epigenetic changes may predispose to neurological disease at a later time, based on DNA methylation and oxidative status of certain gene promoters [22,23,34]. Fetal alcohol exposure may also affect DNA methylation patterns and ensuing disease [35]. Lung diseases are particularly associated with exposure to airborne environmental agents, including cigarette smoke and traffic-related air pollution such as diesel exhaust. These exposures are implicated in diseases like asthma, chronic obstructive pulmonary disease (COPD), emphysema and lung cancer. Exposures can occur *in utero*, in childhood or in adulthood, and the age at which exposures occur may affect the extent of epigenetic changes. The fact that some tissues, including lung and reproductive tissue, develop throughout adolescence and are affected by environmental influences, underscores the importance not only of determining the epigenomes of adult tissues, but also those of cells from embryos and children.

One very interesting example of environmental effects on the epigenome is the population-based analysis of DNA methylation patterns in smokers versus nonsmokers [36,37]. DNA methylation has been favored as the investigated epigenetic mark because it can be determined from purified DNA, which is often the only available biological material in epidemiological studies.

However, instead of DNA obtained from lung cells, it is usually derived from blood cells that are frequently used as a surrogate; in many studies is too costly to obtain, store, and analyze multiple tissue types. Using white blood cell DNA, modest but reproducible DNA methylation changes have been observed in numerous studies [36]. Some of these changes are rapidly reversible upon cessation of exposure, and others can persist for many years, as determined in studies from children exposed in utero and adults examined at different time points following smoking cessation [38,39]. Since the blood cells from which DNA was isolated were not usually sorted by cell type, an important question is whether the observed DNA methylation changes could largely be related to changes in the relative cell type composition of the white blood cells, as recently observed [40]. Corrections for cell type composition are thus required [18,19]. Another key question is whether epigenetic changes seen in the blood reflect changes and disease risk in the lungs. We have recently investigated smoking-associated DNA methylation changes using samples of histologically normal lung from lung cancer patients (unpublished observations). Among others, we detected hypomethylation of cg05575921, which lies in an *AHRR* intron. This same hypomethylation event has also been observed in numerous studies of the white blood cell DNA of smokers [36]. Because we were not able to analyze purified lung cell types we do not know which cells contributed to the observed DNA methylation changes. However, we were able to exclude the possibility that hypomethylation merely reflects changes in tissue composition (such as immune cell infiltration) rather than epigenetic alterations because we carried out functional studies. This illustrates the importance of measuring environmental responses in individual cell types and having clearly established baseline epigenomes in such cells.

Interpretations of the effects of genetic polymorphisms on disease risk requires knowledge of epigenetic marks in the risk-associated cell types

Genetic variants can affect the epigenome in many ways [41–43]. For example, variants in the coding regions of genes involved in establishing or maintaining epigenetic features, such as DNA methyltransferases [44–48] or histone-modifying enzymes [49] have been reported to affect the epigenome on a broad scale, and are frequently detected as somatic mutations in tumors [50–52]. Germline variants have been reported to affect the epigenome at discrete loci in more subtle ways [43,53–55].

Genome-wide association studies (GWAS) performed with microarrays containing millions of probes

for SNPs have identified over 8000 variants associated with human illness [56,57]. Mechanistically connecting GWAS SNPs to disease processes is complicated by two main factors [56,58]:

- Most are co-inherited with hundreds of variants in linkage disequilibrium (LD) that could be more causal than their representative or ‘index’ SNP on the array; and
- Most GWAS and associated LD SNPs lie outside of coding regions of genes, making hypotheses about their function more elusive than for coding SNPs.

Several recent in-depth follow-up studies support the notion that functional noncoding SNPs reside in regulatory regions of the genome, such as gene enhancers, which orchestrate transcriptional programs underpinning health and disease. In all of these studies, the epigenomic features of disease-relevant cell types were profiled and integrated over the locations of GWAS and LD SNPs. This resulted in a shortlist of potential SNP candidates to prioritize in functional follow-up experiments that range from allele-specific ChIP and reporter gene assays [53,54] to genome-wide expression studies following CRISPR-driven deletion of enhancers overlapping risk SNPs [55]. To date, noncoding functional SNPs have been reported to disrupt transcription factor binding sites and enhancer-promoter interactions [53,55], affect DNA methylation [43,59], modulate miRNA and lncRNAs expression and binding [60,61] and affect CTCF occupancy [62].

The importance of evaluating GWAS SNPs in disease-relevant epigenomic contexts is evident in disorders of the respiratory tract. The lungs (as discussed in more detail in the section below) consist of many epithelial cell types that can be broadly categorized as ‘airway’ or ‘alveolar.’ These broad categories harbor subgroups of highly specialized cell types with dramatically distinct phenotypes, epigenomes and disease predispositions [1,63–68]. For example, epithelial cell types lining the conducting airways and bronchioles are implicated in squamous cell lung cancer (SQLC) and disorders related to bronchitis [27,28,69]. Conversely, AT2 cells of the alveolar compartment are thought to give rise both to lung adenocarcinoma (LUAD), the major subtype of lung cancer, and idiopathic pulmonary fibrosis (IPF), a deadly illness in which the alveoli are destroyed and replaced by scar tissue [70,71]. Interestingly, the C-allele of rs2736100, located in an intron of the telomerase-encoding *TERT* gene, is linked to increased risk of LUAD but not SQLC or other lung cancer subtypes in GWA studies [72,73]. In contrast, the rs2736100-A allele is

the only risk SNP that replicates in GWA studies of IPF [74]. Heightened telomerase activity is a well-documented phenomenon in tumor predisposition [75], while decreased activity is linked to epithelial stem cell exhaustion and tissue remodeling seen in emphysema and IPF [76,77]. Not surprisingly, the rs2736100-C allele is associated with elevated *TERT* expression in lung tissue [73], while the rs2736100-A allele is associated with shortened telomeres that are a hallmark of IPF [78–80]. Epigenomic aberrations have been associated with LUAD and IPF at disparate loci in several reports [81–83]. Animal models of both disorders exist [84,85], and epigenomically profiling labeled AT2 cells at different stages of disease progression might provide actionable insight into the mechanism whereby rs2736100 deregulates telomerase. Likewise, epigenomic profiles may elucidate the role of other causal SNPs elsewhere in the genome.

The complexities of determining comprehensive organ-specific epigenomes: the lung as an example

To address the challenges described in the above sections, the epigenomes of all cell types in the human body would need to be mapped. Ideally, this should be performed by experts in cell purification in an organ-by-organ fashion. In this section, we provide a brief overview of the types of cells found in the lung to illustrate the complexity of dissecting the epigenomes of cells comprising whole organs. In the lungs, there are over 40 different types of cells ranging from a variety of epithelial cells to smooth muscle cells, cells composing the vasculature and lymphatic system and immune cells defending the airways against foreign particulate matter and pathogens (Figure 2) [1,63–68]. As outlined in the sections of this review, the individual epigenomes of the different cell types could be used for numerous purposes, including comparison to regenerated tissues and diseased cells, to identify the consequences of environmental exposures and to gain insights into the effects of genetic diversity (Figure 2, right side). An example illustrating the importance of characterizing cell type-specific epigenomes is given in the bottom panel of Figure 2, which compares ChIP-seq data for histone 3 lysine 27 acetylation (H3K27Ac, a mark indicating an active enhancer) in small airway epithelium, normal human lung fibroblasts and unpurified adult lung tissue. It indicates the presence of an enhancer downstream of the *FHAD* gene in small airway epithelial cells but not in lung fibroblasts; the element is not clearly detectable in adult lung, which is a mix of many different cell types. This point is further underscored by studies of two epithelial populations in human peripheral lung, AT1 and AT2 cells; dramatic differ-

ences can be observed between their epigenomes even though they are derived from two cell types linked by a direct progenitor-progeny relationship and co-existing in close proximity in the same organ compartment [86].

The airways begin with the trachea, which connects the throat to the primary bronchi. The latter start at the bifurcation to both lungs and branch out into the secondary bronchi connecting to the lung lobes (three on the right and two on the left in humans due to the presence of the heart). This is followed by the tertiary bronchi, which transport the air into lung segments. The bronchioles follow, transitioning into the terminal bronchioles, considered the smallest parts of the airways. The transition to the alveoli occurs through the respiratory bronchioles. Cell types found in the airways play numerous roles, and are often divided into categories based on their function [63–68]. Basal cells lie on the basement membrane and are thought to be involved in airway regeneration upon injury. They are also implicated as progenitors of squamous cell carcinoma [28]. A variety of secretory cells defend against pathogens and offer protection against contaminants. These include: serous cells, which produce lysozyme, albumin and antileukoprotease, and are implicated in lung repair upon injury; goblet cells, which secrete mucus and have numerous membrane-bound secretory granules; and club cells, which express detoxifying enzymes, the characteristic protein uteroglobin (CC10) and secrete surfactant. Other epithelial cell types found in the airways include ciliated cells, whose function is to move mucus up and out of the airways, thereby ridding the lungs of particulate matter and pathogens, and chemosensory cells, including pulmonary neuroendocrine cells (PNECs) and the pear-shaped brush cells. The latter two cell types are rare, each accounting for less than 1% of lung epithelial cells. PNECs are thought to function in oxygen sensing and as mentioned have been very strongly implicated as the cell of origin of SCLC [29]. The function of brush cells remains enigmatic. It is interesting that they express proteins of the taste transduction pathways, and this may be related to the sensing of pathogens [87]. In addition to the epithelial cells listed above, the airways are lined with smooth muscle, which allow the bronchioles in particular to expand and contract in response to exercise or inhaled particulates respectively. Connective tissue and the vascular and lymphatic systems surround the airways. Lastly, immune cells populate the airways, consisting of specialized macrophages and dendritic cells. The relative cell composition in the upper and lower airways differs, and it has been suggested that even single cell types like basal cells might differ phenotypically dependent on their proximal or distal location [64].

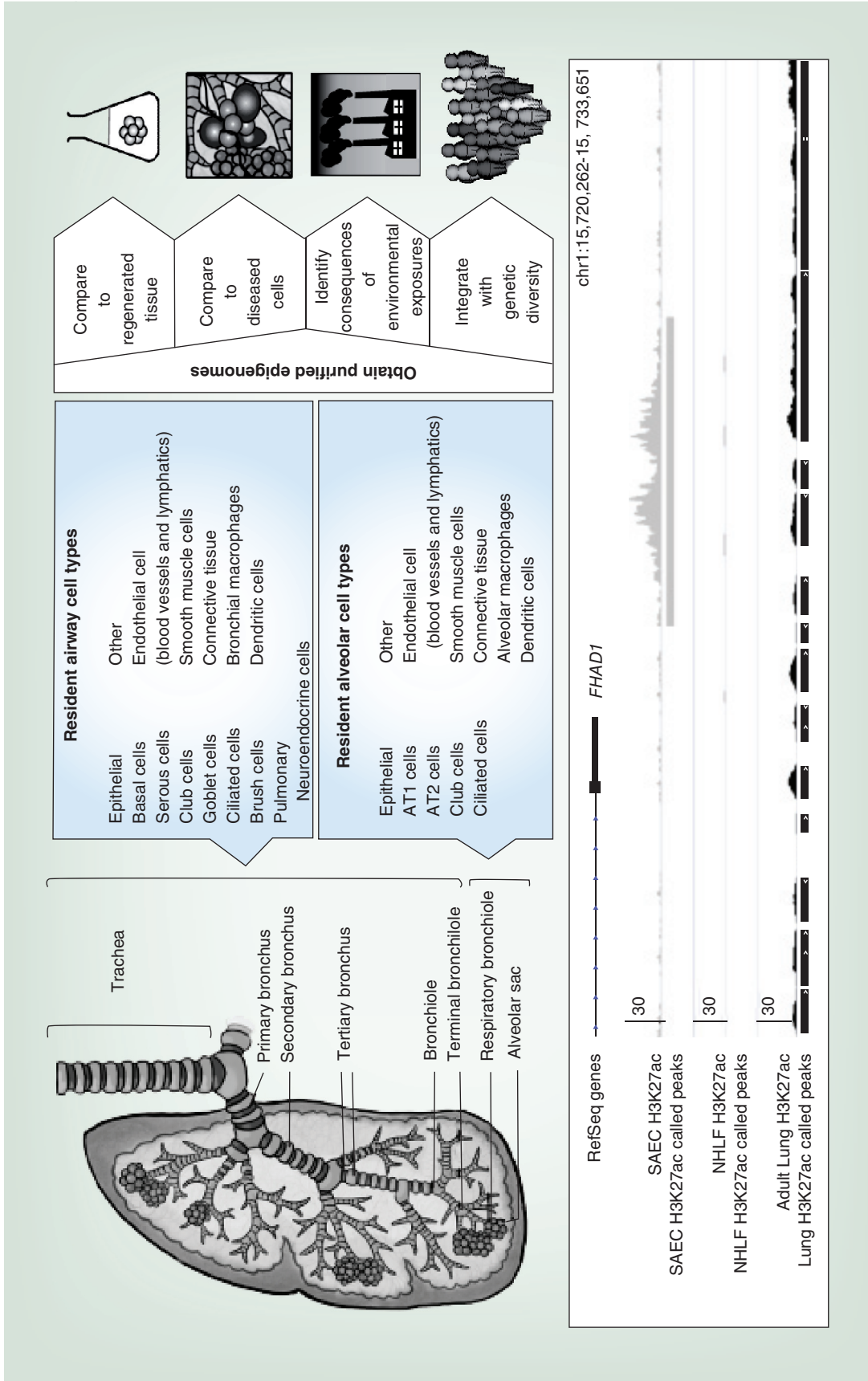


Figure 2. Diagram of the structure of the lung, resident cell types and the uses of cell type-specific epigenomes. Only the right lung is shown. Airway and alveolar compartments and their cell types are indicated. The composition of human respiratory bronchioles has been poorly investigated but is thought to contain club cells and rare ciliated cells. At right, different applications for obtained epigenomes are indicated. At the bottom, an example is shown of epigenetic differences between lung cell types: histone 3 lysine 27 acetylation (H3K27Ac, a mark indicating an active enhancer) ChIP-seq data is compared between small airway epithelium, normal human lung fibroblasts and adult lung tissue. The presence of an enhancer is noted downstream of the *FHAD* gene in small airway epithelial cells but not in lung fibroblasts; the element is not clearly detectable in adult lung, which is a mix of many different cell types.

The alveolar compartment has fewer epithelial cell types. The cellular composition of the human respiratory bronchiole has been poorly characterized but is thought to include club cells and rare ciliated cells. The large delicate AT1 cells constitute the respiratory surface (approximately 95% of the alveolar surface area) and are in intimate contact with the capillary network, while cuboidal AT2 cells produce surface tension-reducing surfactants that prevent the alveolar sacs from collapsing upon exhalation. In addition, AT2 cells are thought to play a role in peripheral lung regeneration upon injury and can transdifferentiate into AT1 cells [88]. This process can be recapitulated *in vitro* over the course of 6–8 days and was used by us to obtain human AT1 cells for epigenomic profiling because the fragile primary human AT1 cells are very difficult to purify in large numbers [86,89]. AT2 cells are the presumed precursors of lung adenocarcinoma [70]. As with the airways, numerous other cells are present: interstitial cells (fibroblasts), a large number of vascular and lymphatic cells, limited smooth muscle, and alveolar macrophages and dendritic cells. The latter are distinct from other macrophages in the body and should be independently studied [90].

From these observations it is clear that samples of lung tissue will contain many different cell types. How much of each type is present in a piece of tissue depends on the sampled location and the natural abundance of each cell type. In an analysis of cell type numbers in an adult human lung, endothelial cells were most numerous (~140 billion), followed by AT2 cells (~70 billion), AT1 cells (~40 billion) and alveolar macrophages (~30 billion) [1]. All airway cells are at least an order of magnitude less common, with the chemosensory cells in lowest abundance.

In order to determine the epigenomes for all the cell types in the lung, each must be separately purified and characterized, a process that can be challenging for delicate cells or cells of relatively low abundance. How this can be approached will be discussed in the next section. Given the technical difficulty and expense of such purifications, they do not lend themselves to be carried out for numerous subjects, for example, for epidemiological studies. Rather, the goal would be to use purified epigenomes from a limited number of individual lungs to identify regulatory elements in regions that may be implicated based on epidemiological studies.

New approaches to obtaining cell-type-specific epigenomes

When attempting to characterize purified human epigenomes, it is of course desirable to obtain primary cells, although for some fragile cell types such as AT1 cells, this can be challenging. In initial epigenomic studies

of the peripheral lung, low-cell-number methods for epigenomic mapping were not yet available, requiring the purification tens of millions of cells to capture high quality data for each histone mark by chromatin immunoprecipitation-high throughput sequencing (ChIP-seq) [86]. Such cell quantities can only be obtained from a deceased person. Because remnant transplant lung was used, the donors had often been on life support for several days. Despite the unnatural conditions, lungs that have been maintained in this way are successfully transplanted into patients, indicating that lung function is sufficiently maintained. For our experiments, remnant transplant lung that was disease-free was transported by air to the laboratory where it was lavaged to remove macrophages. These can be saved for independent profiling. The lung was then digested with elastase, using a procedure adapted from the Matthey laboratory [86,91]. The tissue was subsequently chopped, filtered and subjected to gradient centrifugation. Any remaining immune cells were then removed by panning on IgG-coated plates and AT2 cells were purified on epithelial cell adhesion molecule-carrying microbeads. This yielded very large numbers of AT2 cells that were approximately 90% pure. Because AT1 cells are too fragile to purify in large quantities, their *in vitro* equivalent was obtained by plating AT2 cells and allowing them to spontaneously differentiate [86,89]. Dramatic transcriptional changes occurred within two days of plating and by 6 days the cells had attained an AT1 cell-like phenotype and had formed tight junctions as measured by transepithelial resistance. Culturing cells to obtain derivative cell types introduces potential variables, but it may be the only way large numbers of some cells can be obtained. This type of approach has been successfully used to grow cells from organs such as the kidney, brain and colon with maintenance of the proper cell characteristics [92–95]. In the case of peripheral lung, it allowed us to profile AT2 cells and AT1-like cells for DNA methylation, numerous histone marks and open chromatin, the latter by formaldehyde-assisted identification of regulatory elements (FAIRE)-seq, which allows one to use remnant material from the ChIP experiments. However, dealing with the large number of cells that was needed limited the ability to purify additional cell types simultaneously.

Fortunately, important progress has been made, not only in high throughput sequencing technologies but also in low-cell-number approaches. This means that FACS, which is too cumbersome and expensive to use for obtaining very large numbers of cells, is now a viable option for cell purification. Combined with the development of new low-cell-number methods for epigenomic profiling such as ATAC-seq to map accessible DNA (equivalent to DNase hypersensitive sites) [4], and iChIP-seq [7] or

recovery via protection (RP)-ChIP-seq and favored amplification RP-ChIP-seq (FARP-ChIP-seq) [8,96], it will now be possible to carry out epigenomic profiling using limited numbers of many different kinds of cells. In fact, as technologies progress to single-cell approaches, cell numbers will be further reduced. A single-cell method for ATAC-seq has just been described [96]. However, single cell methods always require a group of similar cells to obtain average signatures and this is expensive and computationally intense.

One important limitation of bead-based or FACS-based cell purification of distinct cell types is that antibodies to extracellular targets must be available to identify the desired cell types. Such antibodies are not in hand for many of the types of cells in the human body. This can be overcome by identifying cell-type specific markers for each cell type. To this end, single cell RNA-seq data for cell populations from all over the body would be extremely valuable [3]. Principal component analysis of single cell populations will allow groups of cells with similar phenotypes to be identified. Such studies are being carried out for many tissues, including the lung and brain [3,97], and will help clarify whether similar cell types from different organs or from different regions of the same organ are distinct, such as epithelial cells found in the bronchi versus the terminal bronchioles, or airway versus alveolar macrophages [98]. Single cell and low cell number techniques will also enable the epigenomic analysis of cells and tissues from living humans through biopsies or different types of noninvasive sampling. This will allow a closer approximation of the natural state of tissues.

What about model organisms? Would it not be much easier to use genetically engineered animals with cell type-specifically labeled nuclei to investigate the composite epigenomes of organs, and then to translate those findings to humans? Elegant procedures have been developed that allow nuclei to be labeled using animals engineered with cell type-specific promoters driving a nuclear envelope protein [99,100]. The tissues of interest can quickly be shredded to purify the labeled nuclei. Indeed, these model organism systems are very valuable, and can yield much information on the epigenetic encoding of tissues, even from organisms distantly related to humans like flies and worms [17]. However, there are numerous differences even between human organs and their mammalian counterparts, and these differences will likely affect cellular epigenomes. In the lung, for example, the types of cells in different compartments can differ markedly between species [63–68].

Future perspective

The epigenomic profiling of cell types within organs, and ultimately of all the cell types in the human body,

will provide key information supporting biomedical investigations. In the short term, the focus should be placed on elucidating the collective epigenomes for a limited number of adult individuals of both genders. However, such studies would form just the very beginning of whole organism epigenomic analyses. We must build upon them by determining the comparative epigenomes from individuals of different ages, including embryos, infants, children and adolescents, different ethnicities and a wide range of environmental exposures. Only then will we truly begin to grasp the range of epigenetic variation within and between individuals. Cost and complexity would prevent the examination of the large numbers of individuals usually enrolled in epidemiological studies. Instead, a limited number of epigenomes would be used to provide a basis for the investigation of very specific genomic regions or epigenetic marks in a large number of individuals. We anticipate that single-cell techniques will further evolve to allow the analysis of all epigenetic marks of interest in single cells, and that sequencing approaches will improve to enable the determination of numerous epigenetic marks from each individual cell. This will, in turn, permit real-time sampling of dynamic epigenomes in patients to inform personalized medical strategies aimed at determination of disease risk, prevention and treatment.

Acknowledgements

The authors acknowledge AN Laird and P Marconett for help with illustrations in **Figure 2**.

Disclaimer

None of the funding agencies had any role in the content or timing of the publication of this manuscript; the content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Financial & competing interests disclosure

The authors were supported by the Hastings and Whittier Foundations, donations from C Pembroke and W Pembroke, J Glick, L Auerbach, NE Warner, HJ Norris, and the Canary and Thomas G. Labrecque Foundations to IA Laird-Offringa, Department of Defense Concept Award (W81XWH1410174) to IA Laird-Offringa, NIH grant (R01 HL114094) to IA Laird-Offringa and Z Borok, NIH grants (P30 H101258, R01HL112638 and 4 R37HL062569) to Z Borok and (1R01HL114959) to B Zhou, and the Norris Comprehensive Cancer Center core grant P30CA0189 from the National Cancer Institute. TR Stueve was supported by the National Institute of Environmental Health Sciences (NIEHS, NIH T32ES013678) and by the USC Provost's Postdoctoral Scholar Research Grant. CN Marconett was supported by an ACS/Canary postdoctoral fellowship number PFTED-10-207-01-SIED, R01 HL114094 and the Department of Surgery. Z Borok is the

Edgington Chair in Medicine. CN Marconett and IA Laird-Offringa are inventors listed on pending USC patent application 61/817,840, entitled "Accurate in Vitro Copying of DNA Methylation". The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Executive summary

Examination of purified cell types is crucial for tissue regeneration research

- Epigenomic profiles are required to verify the integrity of stem cells and their derivatives used for cell-based therapies and tissue regeneration.

Identifying disease-associated epigenetic aberrations will require knowledge of progenitor cells & establishment of reference epigenomes

- Within a tissue distinct cell types can be affected by a variety of diseases.
- In order to establish how the epigenomes of diseased cells have been affected, the baseline epigenome of the cell of origin for each particular disease must be established.

Measuring the effects of environmental exposures requires knowledge of the baseline epigenomes of unexposed cell types

- Environmental exposures can be measured in disease-prone and surrogate tissues.
- Determination of epigenetic changes can inform prevention and therapeutic strategies.

Interpretations of the effects of genetic polymorphisms on disease risk requires knowledge of epigenetic marks in the risk-associated cell types

- Most SNPs lie in introns or intergenic regions.
- The biological function of SNPs, therefore, depends on whether they disrupt epigenetic regulatory elements in disease-associated cell types.

The complexities of determining comprehensive organ-specific epigenomes: the lung as an example

- Each organ consists of a wide variety of different cell types with distinct functions and disease predispositions.
- The epigenomes of closely related cells in the same organ compartment can differ dramatically.
- Different cell types must be individually purified and epigenomically characterized.

New approaches to obtaining cell-type-specific epigenomes

- FACS sorting and culture of purified cell types can provide cells for epigenomic analyses.
- Low-cell-number and single cell techniques are lowering the numbers of cells needed for epigenomic profiling, facilitating analyses.

Future perspective

- Efficient single-cell epigenome analysis will allow a personalized approach to assessment of genetic risk, environmental exposure and disease states.

References

Papers of special note have been highlighted as:

• of interest; •• of considerable interest

- 1 Bianconi E, Piovesan A, Facchin F *et al.* An estimation of the number of cells in the human body. *Ann. Hum. Biol.* 40, 463–471 (2013).
- 2 Alberts B, Johnson A, Lewis J *et al.* *Molecular Biology of The Cell (6th Edition)*. Garland Science, Hamden, CT, USA (2014).
- 3 Wu AR, Neff NF, Kalisky T *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* 11, 41–46 (2014).
- 4 Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.1–9 (2015).
- 5 Smallwood SA, Lee HJ, Angermueller C *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820 (2014).
- 6 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330 (2015).
- 7 Lara-Astiaso D, Weiner A, Lorenzo-Vivas E *et al.* Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949 (2014).
- 8 Zheng X, Yue S, Chen H, Weber B, Jia J, Zheng Y. Low-cell-number epigenome profiling aids the study of lens aging and hematopoiesis. *Cell Rep.* 13, 1505–1518 (2015).

- 9 Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. *Science* 330, 612–616 (2010).
- 10 Downen JM, Fan ZP, Hnisz D *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387 (2014).
- 11 Globisch D, Münzel M, Müller M *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* 5, e15367 (2010).
- 12 Simonson OE, Domogatskaya A, Volchkov P, Rodin S. The safety of human pluripotent stem cells in clinical treatment. *Ann. Med.* 47, 370–380 (2015).
- 13 Werbowetski-Ogilvie TE, Bossé M, Stewart M *et al.* Characterization of human embryonic stem cells with features of neoplastic progression. *Nat. Biotechnol.* 27, 91–97 (2009).
- 14 Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* 15, R31 (2014).
- Discusses the importance of using purified populations of cells for epigenetic studies.
- 15 Bataille A, Galichon P, Ziliotis MJ, Sadia I, Hertig A. Epigenetic changes during sepsis: on your marks. *Crit. Care* 19, 358 (2015).
- 16 Maze I, Shen L, Zhang B *et al.* Analytical tools and current challenges in the modern era of neuroepigenomics. *Nat. Neurosci.* 17, 1476–1490 (2014).
- 17 Brown JB, Celniker SE. Lessons from modENCODE. *Annu. Rev. Genomics. Hum. Genet.* 16, 31–53 (2015).
- Provides an overview of model organisms data from the modENCODE project.
- 18 Houseman EA, Accomando WP, Koestler DC *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13, 86 (2012).
- 19 Houseman EA, Kim S, Kelsey KT, Wiencke JK. DNA methylation in whole blood: uses and challenges. *Curr. Environ. Health. Rep.* 2, 145–154 (2015).
- 20 Jimenez-Chillaron JC, Ramon-Krauel M, Ribo S, Diaz R. Transgenerational epigenetic inheritance of diabetes risk as a consequence of early nutritional imbalances. *Proc. Nutr. Soc.* 75, 78–89 (2016).
- 21 Jiménez-Chillaron JC, Díaz R, Martínez D *et al.* The role of nutrition on epigenetic modifications and their implications on health. *Biochimie* 94, 2242–2263 (2012).
- 22 Lahiri DK, Maloney B, Zawia NH. The LEARN model: an epigenetic explanation for idiopathic neurobiological diseases. *Mol. Psychiatry* 14, 992–1003 (2009).
- 23 Lahiri DK, Maloney B. The “LEARN” (Latent Early-life Associated Regulation) model integrates environmental risk factors and the developmental basis of Alzheimer’s disease, and proposes remedial steps. *Exp. Gerontol.* 45, 291–296, (2010).
- 24 Maze I, Wenderski W, Noh KM *et al.* Critical role of histone turnover in neuronal transcription and plasticity. *Neuron* 87, 77–94 (2015).
- 25 Nestler EJ, Peña CJ, Kundakovic M, Mitchell A, Akbarian S. Epigenetic basis of mental illness. *Neuroscientist* doi:10.1177/1073858415608147 (2016) (Epub ahead of print).
- 26 Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications: *Nat. Rev. Cancer* 11, 726–734 (2011).
- 27 Hanna JM, Onaitis MW. Cell of origin of lung cancer. *J. Carcinog.* 12, 6 (2013).
- 28 Sutherland KD, Berns A. Cell of origin of lung cancer. *Mol. Oncol.* 4, 397–403 (2010).
- 29 Sutherland KD, Proost N, Brouns I, Adriaensen D, Song JY, Berns A. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* 19, 754–764 (2011).
- 30 Kalari S, Jung M, Kernstine KH, Takahashi T, Pfeifer GP. The DNA methylation landscape of small cell lung cancer suggests a differentiation defect of neuroendocrine cells. *Oncogene* 32, 3559–3568 (2013).
- 31 Poirier JT, Gardner EE, Connis N *et al.* DNA methylation in small cell lung cancer defines distinct disease subtypes and correlates with high expression of *EZH2*. *Oncogene* 34, 5869–5878 (2015).
- 32 Fraga MF, Ballestar E, Paz MF *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* 102, 10604–10609 (2005).
- Highlights the key role analysis of mono- and dizygotic twins plays in determining genetic versus epigenetic effects.
- 33 Busche S, Shao X, Caron M *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).
- 34 Bennett DA, Yu L, Yang J, Srivastava GP, Aubin C, De Jager PL. Epigenomics of Alzheimer’s disease. *Transl. Res.* 165, 200–220 (2015).
- 35 Resendiz M, Chen Y, Oztürk NC, Zhou FC. Epigenetic medicine and fetal alcohol spectrum disorders. *Epigenomics* 5, 73–86 (2013).
- 36 Gao X, Jia M, Zhang Y, Breitling LP, Brenner H. DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. *Clin. Epigenetics* 7, 113 (2015).
- 37 Shenker NS, Polidoro S, van Veldhoven K *et al.* Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Hum. Mol. Genet.* 22, 843–851 (2013).
- 38 Tsaprouni LG, Yang TP, Bell J *et al.* Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 9, 1382–1396 (2014).
- 39 Novakovic B, Ryan J, Pereira N, Boughton B, Craig JM, Saffery R. Postnatal stability, tissue, and time specific effects of AHRR methylation change in response to maternal smoking in pregnancy. *Epigenetics* 9, 377–386 (2014).
- 40 Bauer M, Linsel G, Fink B *et al.* A varying T cell subtype explains apparent tobacco smoking induced single CpG hypomethylation in whole blood. *Clin. Epigenetics* 7, 81 (2015).
- 41 Grundberg E, Meduri E, Sandling JK *et al.* Global analysis of DNA methylation variation in adipose tissue from twins

- reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* 93, 876–890 (2013).
- 42 Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* 10, 1024–1032 (2015).
- 43 Shi J, Marconett CN, Duan J *et al.* Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat. Commun.* 5, 3365 (2014).
- 44 Arakawa Y, Watanabe M, Inoue N, Sarumaru M, Hidaka Y, Iwatani Y. Association of polymorphisms in *DNMT1*, *DNMT3A*, *DNMT3B*, *MTHFR* and *MTRR* genes with global DNA methylation levels and prognosis of autoimmune thyroid disease. *Clin. Exp. Immunol.* 170, 194–201 (2012).
- 45 Kullmann K, Deryal M, Ong MF, Schmidt W, Mahlknecht U. *DNMT1* genetic polymorphisms affect breast cancer risk in the central European Caucasian population. *Clin. Epigenetics* 5, 7 (2013).
- 46 Saradalekshmi KR, Neetha NV, Sathyan S, Nair IV, Nair CM, Banerjee M. DNA methyl transferase (*DNMT*) gene polymorphisms could be a primary event in epigenetic susceptibility to schizophrenia. *PLoS ONE* 9, e98182 (2014).
- 47 Wu H, Zhang K, Gong P *et al.* A novel functional TagSNP Rs7560488 in the *DNMT3A1* promoter is associated with susceptibility to gastric cancer by modulating promoter activity. *PLoS ONE* 9, e92911 (2014).
- 48 Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. *Epigenomics* 7, 247–265 (2015).
- 49 Cebrian A, Pharoah PD, Ahmed S *et al.* Genetic variants in epigenetic genes and breast cancer risk. *Carcinogenesis* 27, 1661–1669 (2006).
- 50 Gonzalez-Perez A, Jene-Sanz A, Lopez-Bigas N. The mutational landscape of chromatin regulatory factors across 4,623 tumor samples. *Genome Biol.* 14, r106 (2013).
- 51 Weirich S, Kudithipudi S, Kycia I, Jeltsch A. Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. *Clin. Epigenetics* 7, 36 (2015).
- 52 Scourzic L, Mouly E, Bernard OA. TET proteins and the control of cytosine demethylation in cancer. *Genome Med.* 7, 9 (2015).
- 53 Huang Q, Whittington T, Gao P *et al.* A prostate cancer susceptibility allele at 6q22 increases *RFX6* expression by modulating *HOXB13* chromatin binding. *Nat. Genet.* 46, 126–135 (2014).
- 54 Rhie SK, Coetzee SG, Noushmehr H *et al.* Comprehensive functional annotation of seventy-one breast cancer risk Loci. *PLoS ONE* 8, e63925 (2013).
- 55 Yao L, Tak YG, Berman BP, Farnham PJ. Functional annotation of colon cancer risk SNPs. *Nat. Commun.* 5, 5114 (2014).
- 56 Coetzee SG, Rhie SK, Berman BP, Coetzee GA, Noushmehr H. FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs. *Nucleic Acids Res.* 40, e139 (2012).
- 57 Welter D, MacArthur J, Morales J *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006 (2014).
- 58 Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* 8, 57 (2015).
- **Discusses the use of epigenomic profiles to identify functional risk SNPs.**
- 59 Liu Y, Li X, Aryee MJ *et al.* GeMes, clusters of DNA methylation under genetic control, can inform genetic and epigenetic analysis of disease. *Am. J. Hum. Genet.* 94, 485–495 (2014).
- 60 Jendrzewski J, He H, Radomska HS *et al.* The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc. Natl Acad. Sci. USA* 109, 8646–8651 (2012).
- 61 Hazelett DJ, Rhie SK, Gaddis M *et al.* Comprehensive functional annotation of 77 prostate cancer risk loci. *PLoS Genet.* 10, e1004102 (2014).
- 62 Dai J, Zhu M, Wang C *et al.* Systematical analyses of variants in CTCF-binding sites identified a novel lung cancer susceptibility locus among Chinese population. *Sci. Rep.* 5, 7833 (2015).
- 63 Reynolds SD, Pinkerton KE, Mariassy AT. Epithelial cells of trachea and bronchi. In: *Comparative Biology of the Normal Lung*. Parent RA (Ed). Academic Press, San Diego, USA, 61–81 (2015).
- 64 Rackley CR, Stripp BR. Building and maintaining the epithelium of the lung. *J. Clin. Invest.* 122, 2724–2730 (2012).
- 65 Mercer RR, Crapo JD. Architecture of the gas exchange region of the lungs. In: *Comparative Biology of the Normal Lung (Second Edition)*. Parent RA, (Ed.), Academic Press, San Diego, USA, 93–104 (2015).
- 66 Pinkerton KE, Van Winkle LS, Plopper CG, Smiley-Jewell S, Covarrubias EC, McBride JT. Architecture of the Tracheobronchial Tree. In: *Comparative Biology of the Normal Lung (Second Edition)*. Parent RA, (Ed.), Academic Press, San Diego, USA, 33–51 (2015).
- 67 Pinkerton KE, Van Winkle LS, Plopper CG. Overview of diversity in the respiratory system of mammals. In: *Comparative Biology of the Normal Lung (Second Edition)*. Parent RA, (Ed.), Academic Press, San Diego, USA, 3–5 (2015).
- 68 Plopper CG, Hyde DM. Epithelial Cells of the Bronchiole. In: *Comparative Biology of the Normal Lung (Second Edition)*. Parent RA, (Ed.), Academic Press, San Diego, USA, 83–92 (2015).
- 69 Thurlbeck WM. The pathology of small airways in chronic airflow limitation. *Eur. J. Respir. Dis. Suppl.* 121, 9–18 (1982).
- 70 Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* 507, 190–194 (2014).
- 71 Willis BC, Liebler JM, Luby-Phelps K *et al.* Induction of epithelial-mesenchymal transition in alveolar epithelial

- cells by transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. *Am. J. Pathol.* 166, 1321–1332 (2005).
- 72 Landi MT, Chatterjee N, Yu K *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Hum. Genet.* 85, 679–691 (2009).
- 73 Wei R, Cao L, Pu H *et al.* TERT polymorphism rs2736100-C is associated with EGFR mutation-positive non-small cell lung cancer. *Clin. Cancer Res.* 21, 5173–5180 (2015).
- 74 Fingerlin TE, Murphy E, Zhang W *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat. Genet.* 45, 613–620 (2013).
- 75 Kim NW, Piatyszek MA, Prowse KR *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011–2015 (1994).
- 76 Hao LY, Armanios M, Strong MA *et al.* Short telomeres, even in the presence of telomerase, limit tissue renewal capacity. *Cell* 123, 1121–1131 (2005).
- 77 Armanios M, Blackburn EH. The telomere syndromes. *Nat. Rev. Genet.* 13, 693–704 (2012).
- 78 Codd V, Nelson CP, Albrecht E *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* 45, 422–427, 427e1 (2013).
- 79 Walsh KM, Codd V, Smirnov IV *et al.* Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nat. Genet.* 46, 731–735 (2014).
- 80 Alder JK, Chen JJ, Lancaster L *et al.* Short telomeres are a risk factor for idiopathic pulmonary fibrosis. *Proc. Natl Acad. Sci. USA* 105, 13051–13056 (2008).
- 81 Suzuki A, Makinoshima H, Wakaguri H *et al.* Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. *Nucleic Acids Res.* 42, 13557–13572 (2014).
- 82 Yang IV, Pedersen BS, Rabinovich E *et al.* Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 190, 1263–1272 (2014).
- 83 Sanders YY, Ambalavanan N, Halloran B *et al.* Altered DNA methylation profile in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 186, 525–535 (2012).
- 84 Kim CF, Jackson EL, Kirsch DG *et al.* Mouse models of human non-small-cell lung cancer: raising the bar. *Cold Spring Harb. Symp. Quant. Biol.* 70, 241–250 (2005).
- 85 B Moore B, Lawson WE, Oury TD, Sisson TH, Raghavendran K, Hogaboam CM. Animal models of fibrotic lung disease. *Am. J. Respir. Cell. Mol. Biol.* 49, 167–179 (2013).
- 86 Marconett CN, Zhou B, Rieger ME *et al.* Integrated transcriptomic and epigenomic analysis of primary human lung epithelial cell differentiation. *PLoS Genet.* 9, e1003513 (2013).
- 87 Reid L, Meyrick B, Antony VB, Chang LY, Crapo JD, Reynolds HY. The mysterious pulmonary brush cell: a cell in search of a function. *Am. J. Respir. Crit. Care Med.* 172, 136–139 (2005).
- 88 Adamson IY, Bowden DH. Derivation of type 1 epithelium from type 2 cells in the developing rat lung. *Lab. Invest.* 32, 736–745 (1975).
- 89 Marconett CN, Zhou B, Siegmund KD, Borok Z, Laird-Offringa IA. Transcriptomic profiling of primary alveolar epithelial cell differentiation in human and rat. *Genom. Data* 2, 105–109 (2014).
- 90 Hussell T, Bell TJ. Alveolar macrophages: plasticity in a tissue-specific context. *Nat. Rev. Immunol.* 14, 81–93 (2014).
- 91 Ballard PL, Lee JW, Fang X *et al.* Regulated gene expression in cultured type II cells of adult human lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 299, L36–50 (2010).
- 92 Helbert MJ, Dauwe SE, De Broe ME. Flow cytometric immunodissection of the human distal tubule and cortical collecting duct system. *Kidney Int.* 59, 554–564 (2001).
- 93 Van der Hauwaert C, Savary G, Gnemmi V *et al.* Isolation and characterization of a primary proximal tubular epithelial cell model from human kidney by CD10/CD13 double labeling. *PLoS ONE* 8, e66750 (2013).
- 94 Rozental R, Gebhard D, Padin C *et al.* Purification of cell populations from human fetal brain using flow cytometric techniques. *Brain Res. Dev. Brain Res.* 85, 161–170 (1995).
- 95 Jung P, Sato T, Merlos-Suárez A *et al.* Isolation and *in vitro* expansion of human colonic stem cells. *Nat. Med.* 17, 1225–1227 (2011).
- 96 Buenrostro JD, Wu B, Litzenburger UM *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490 (2015).
- 97 Darmanis S, Sloan SA, Zhang Y *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* 112, 7285–7290 (2015).
- 98 Van de Laar E, Clifford M, Hasenoeder S *et al.* Cell surface marker profiling of human tracheal basal cells reveals distinct subpopulations, identifies MST1/MSP as a mitogenic signal, and identifies new biomarkers for lung squamous cell carcinomas. *Respir. Res.* 15, 160 (2014).
- 99 Deal RB, Henikoff S. The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat. Protoc.* 6, 56–68 (2011).
- 100 Steiner FA, Talbert PB, Kasinathan S, Deal RB, Henikoff S. Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.* 22, 766–777 (2012).