

Genetic and epigenetic contribution to complex traits

Helena Kilpinen^{1,2,3} and Emmanouil T. Dermitzakis^{1,2,3,*}

¹Department of Genetic Medicine and Development, University of Geneva Medical School, 1 Rue Michel-Servet, Geneva 1211, Switzerland, ²Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva 1211, Switzerland, ³Swiss Institute of Bioinformatics, Geneva 1211, Switzerland

Received August 14, 2012; Revised and Accepted September 6, 2012

Much of the recent advances in functional genomics owe to developments in next-generation sequencing technology, which has contributed to the exponential increase of genomic data available for different human disease and population samples. With functional sequencing assays available to query both the transcriptome and the epigenome, annotation of the non-coding, regulatory genome is steadily improving and providing means to interpret the functional consequences of genetic variants associated with human complex traits. This has highlighted the need to better understand the normal variation in various cellular phenotypes, such as epigenetic modifications, and their transgenerational inheritance. In this review, we discuss different aspects of epigenetic variation in the context of DNA sequence variation and its contribution to complex phenotypes.

INTRODUCTION

Genome-wide genetic information for different human populations and phenotypes is becoming increasingly abundant due to the efforts of projects such as the 1000 Genomes project (1) and the recent wave of genome-wide association studies (GWASs) in human complex traits (2). Due to exceptionally large study samples, thousands of genetic variants, primarily single-nucleotide polymorphisms (SNPs), have been associated with diseases and other trait phenotypes with very good estimates of the effect size and the significance of the effects. While this is a major milestone in human genetics, an important component of these studies, i.e. the understanding and interpretation of the function of the associated variants, is still largely lacking. It has become evident that most of these variants are regulatory rather than coding, which has emphasized the urgent need for better functional annotation of the human genome. To address this, large-scale studies such as the ENCODE project (3) were set out to identify and catalog all functional elements in the human and other mammalian genomes, and while these efforts have provided a tremendous amount of new information, we are still far from understanding the default function and interplay or such elements in different tissues and developmental stages. In this context, trying to evaluate the impact of specific nucleotide

changes on the functionality of such elements is a challenge. Thus, we are still a long way from sequencing a single genome and inferring its deviation from the 'standard' genome or cellular function.

A key method to disentangle the functionality of genetic variants is to directly measure their molecular effects. This can be done by analyzing DNA sequence variation for association with different intermediate phenotypes at the cellular level. This approach helps to reduce the indirect link between a genetic variant and a whole-organism phenotype (such as a disease) to more direct links between these variants and their intermediate effects. There are many assays to study different types of molecular and cellular phenotypes (4), but some of the most popular, well-understood and well-developed ones measure epigenetic modifications, i.e. chemical modifications of DNA or chromatin. Such modifications are often viewed as proxies to regulatory changes occurring at a genomic region and are therefore highly informative of local molecular effects. Epigenetic modifications can result from either genetic variants that influence the epigenetic state of a genomic region or environmental effects that are stored and remembered in the form of epigenetic modifications. In fact, it is now well accepted that much of the memory of our body of past environmental exposures and behavioral effects is mediated by epigenetic changes in the

*To whom correspondence should be addressed. Tel: +41 223795483; Fax +41 223795706; Email: emmanouil.dermitzakis@unige.ch

relevant cell types and tissues (5). This is of great importance as it follows from this that much of the information on human phenotypic variability and disease risk can be accessed directly by assaying appropriate cells from our body. Therefore, understanding of epigenetic variation is crucial to the understanding of both genome function and environmental effects on the individual genomes of our cells.

In this review, we will discuss the technological and methodological advances that have transformed the study of epigenetic variation. We will touch on the genetic and environmental causes of epigenetic variation and their interplay, as well as the recently introduced concept of epigenome-wide association studies (EWASs). We will discuss what such studies can tell us about causality of epigenetic factors, and what they cannot.

TECHNOLOGICAL ADVANCES ARE ENABLING NEW APPROACHES

According to the current view, genetic information carried by the DNA sequence is embedded in a complex and dynamic 'cloud' of chromatin and associated proteins such as transcription factors that define the readout of the DNA sequence. The backbone of this cloud, loosely defined as the epigenome, is made up of nucleosomes, the building blocks of chromatin. Nucleosomes in turn consist of DNA wrapped around histone proteins, which are subject to many different types of covalent modifications (reviewed, for example, in 6) and whose functional consequences are incompletely understood. Methylation of both DNA and RNA molecules is considered an epigenetic mechanism, as well as the silencing effects mediated by the expanding pool of non-coding RNAs of different lengths, which have been shown to regularly interact with chromatin (7). Finally, it will not be sufficient much longer to study the genome simply as a linear structure, as increasing evidence suggests complex three-dimensional interactions between distant genomic loci as well as the genome and the nuclear lamina (8,9). It has recently been shown that such interactions are closely related to different epigenetic domains (10). Understanding the interplay of these organizational layers of our genome is one of the main aims of current epigenomic research.

Histone modifications, which typically affect the tails of the core histone molecules protruding out of the nucleosome, can be divided roughly into activating and repressive modifications, based on their effects on transcription. The number of possible modifications and modifiable sites in humans is huge, and new ones are continuously being discovered. Thus, with research and antibody production lagging behind, our knowledge of the functional role of individual histone modifications remains superficial at best. Studying histone modifications and other epigenetic markers in a combinatorial manner across the genome to define different chromatin states and domains and their interactions with the DNA molecule is likely to be the most powerful approach to elucidate their functions, as has already been demonstrated by several studies (see, for example, 11,12).

Much of the recent advances in epigenomic studies owe to developments in next-generation sequencing (NGS)

technology, which has transformed the type and amount of information that can be extracted with traditional techniques such as chromatin immunoprecipitation or bisulphate sequencing in DNA methylation studies. The NGS platform has also enabled the rapid development of new assays (for example, Gro-seq) (13) and the scale-up of pre-existing methods such as chromosome conformation capture (3C)-based techniques, which can now be performed globally, genome-wide (Hi-C) (14). Significant contributions to method development have also been made by large, collaborative research projects like the ENCODE (3). As the resolution of most functional NGS assays approaches a single base pair, allele-specific readout genome-wide, they have become directly relevant to genetic mapping studies of complex traits, most of which struggle to interpret the functional consequences of the identified DNA variants. Also, importantly, NGS has helped to bring epigenomics as a field conceptually closer to other 'omics' fields, by improving the reproducibility and precision of results.

GENETIC VARIATION AND THE EPIGENOME

The word 'epigenome' has resulted in a number of misconceptions. While the Greek prefix 'epi-' states that the effects or modifications are 'on' the genome, epigenetics has frequently been misunderstood as effects that are independent of the genome. In fact, statements such as 'it is not the genome, it is the epigenome' are not uncommon in the field. Although such classification might appear purely philosophical, it has a profound impact on the approaches taken to study phenotypic variation in humans. Instead of viewing genetics and epigenetics as two conflicting areas that one needs to take sides on, it is much more productive to view them as two highly connected concepts. For example, it is known that epigenetic variation at a given genomic locus can be directly caused by variation in the DNA sequence. On the other hand, epigenetic changes triggered by environmental effects can act as modifiers to genetic variants. In one of the first studies of genetic variation affecting epigenetic properties, family-based cell lines were assayed for DNaseI hypersensitivity (DNaseI HS), a proxy for open chromatin structure, and binding of CCCTC-binding factor (CTCF) (15). The study demonstrated that ~65% of the variation in CTCF binding is heritable due to transmitted DNA sequence variation, and the same trend was observed for DNaseI HS sites. In this example, epigenetic variation can simply be seen as a molecular phenotype.

Epigenetic variation has been proposed as an explanation for the large amount of missing heritability in complex traits (16,17), which has created a lot of confusion. Epigenetic variation can undoubtedly contribute to disease risk and cause phenotypic variability, but for it to contribute to complex trait heritability *independent of* genetic variation one has to assume that variable epigenetic modifications are transmitted from one generation to the next intact, i.e. without DNA variation acting as the determining and transmitting factor. While there is some evidence that such modifications are present in other mammals (18), evidence from human populations is lacking. Such effects in humans are likely to exist, but the inability to detect them to date suggests that they are not

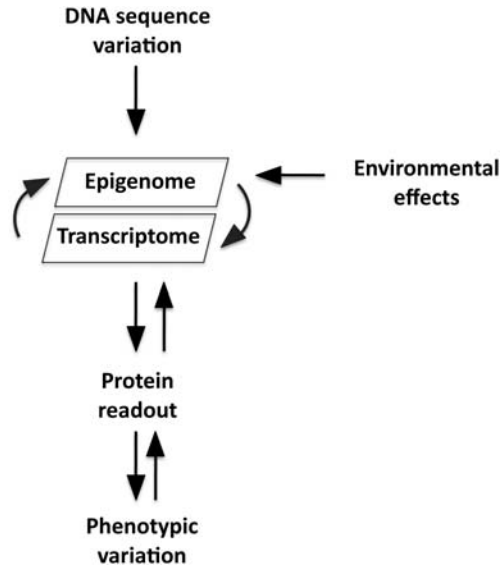


Figure 1. Directionality of molecular effects.

abundant enough to significantly contribute to the heritability of complex traits.

The combination of genetic and epigenetic information can also provide direct links of causality of DNA sequence variation. A recent study assayed DNase I hypersensitivity in 70 cell lines from a population sample and reported thousands of genetic variants affecting the level of chromatin accessibility (DNase I sensitivity quantitative trait loci, dsQTL) (19). They observed that 16% of their dsQTLs affected the expression of nearby genes (i.e. were also classified as expression quantitative trait loci, eQTL), and subsequently estimated that as many as ~55% of eQTLs influence expression variation through changes in chromatin accessibility. Other studies have linked DNA sequence variation to methylation and gene expression in a similar manner (20). While the link between the three measurements seems to some extent easy to derive (for example, SNP to methylation to expression), care should be taken when making conclusions of the directionality of effects (Fig. 1). Contrary to a genetic association where the direction of the effect is by definition from the SNP to the phenotype, an association between an epigenetic mark and a phenotype, in the absence of further information such as longitudinal measurements, provides no evidence as to which one is cause and which is effect. Multi-dimensional studies that link DNA sequence variation, epigenetic modifications, and whole organism phenotypes are, and will be, very important for the detailed characterization of human genetic functional variation. They will also contribute towards the establishment of a genomic blueprint, which will enable the functional and phenotypic interpretation of personalized genomes in the near future.

NON-GENETIC EPIGENETIC VARIATION

The epigenetic layout of the genome is influenced by both genetic and environmental factors. Epigenetic variation can be directly linked to genetic variation (15), and environmental

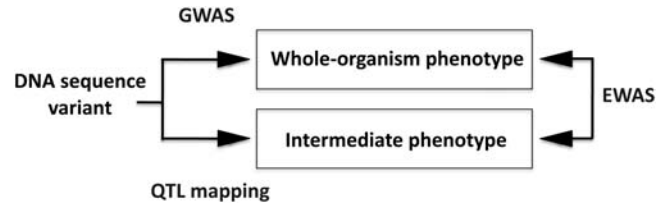


Figure 2. Design of genomic studies. DNA sequence variants can have direct effects on whole-organism phenotypes (e.g. disease status) and/or multiple intermediate, cellular phenotypes (e.g. gene expression, DNA methylation, transcription factor binding and histone modifications). Such effects can be analyzed with GWASs and by mapping QTLs. If association of intermediate phenotypes with the whole-organism phenotype is analyzed (e.g. with EWASs), the distinction between cause and consequence cannot be established without additional information.

effects, such as those experienced *in utero*, can be mediated by epigenetic modifications (5,21). As the heritability of complex phenotypes is never 100% and they thus include a significant non-genetic, potentially epigenetic component, the ability to separate between DNA sequence mediated and non-genetic epigenetic variation will be important. It is also important to identify and distinguish between stable and dynamic epigenetic changes, as dynamic modifications are the most likely proxies for environmental effects.

It has been proposed that epigenomic equivalents of GWASs, i.e. EWASs, could be used to measure the proportion of phenotypic variance between any two individuals that is explained by epigenetic variation (22). In practice, this would mean analyzing suitable epigenetic markers, most plausibly DNA methylation levels, in a large collection of cases and controls for association to a given complex trait (Fig. 2). The biggest challenge of EWASs is likely to be distinguishing between effects mediated by genetic variants versus truly epigenetic properties, as well as the causality versus consequentiality of the identified effects. Thus, an EWAS discovery-stage study sample should ideally consist of monozygotic (MZ) twins, to ensure identification of epigenetic variation independent of the genetic background that could then be replicated in a more traditional case–control sample or, better still, a prospective cohort that would help to distinguish between the cause and consequence of the epigenomic variant (22).

Such approach was taken by Rakyan *et al.* (23) who identified DNA methylation variable positions significantly correlated with type 1 diabetes (T1D) in discordant MZ twins. The authors first confirmed the finding in a separate set of T1D-discordant twins and, using further case–control study samples, were able to show that the identified variants precede disease diagnosis. Such rigorous efforts to distinguish the temporal origin of any trait-associated epigenetic variants will be needed, if the problem of causality is to be resolved. However, for most phenotypes, it will not be trivial to accumulate such study samples, especially if any other biological material than DNA is required.

A few other examples of EWASs have also been published. Breitling *et al.* (24) measured DNA methylation at CpG sites across the genome from 177 current, former and non-smokers and identified a single site, which showed genome-wide significant association with lowered methylation levels in heavy

smokers. The finding was replicated in an independent sample of 328 people. In another study (25), Bell *et al.* studied 172 twins and identified differentially methylated regions associated with age and other aging-related phenotypes. Whilst in the first study the possible influence of genetic variation was not considered, the second study suggested that in a small subset of genes, DNA methylation might mediate genetic association with age-related phenotypes, highlighting the difficulty in interpreting the causality and directionality of epigenetic variation.

Overall, given the dynamic nature of epigenetic variation, it is hard to estimate whether comparisons of epigenetic phenotypes in a case-control setting will be any more useful than such comparisons of gene expression data in complex trait phenotypes. Unlike GWASs, all epigenomic studies face the problem of tissue specificity, and until more tissues relevant to the phenotypes in question become available for analysis in sufficient numbers, the conclusions that can be made from these studies remain limited. Much like gene expression, which as a phenotype is not well suited for an EWAS, analysis of epigenetic variation is likely to be most successful when combined with the analysis of genetic variants (i.e. QTL mapping).

CONCLUSIONS AND FUTURE PERSPECTIVES

In the next few years, our understanding of the multiple layers of genomic information is likely to improve significantly, as new functional sequencing assays will be developed and applied to larger numbers of samples and tissues. Similarly to the data released by the International HapMap Project (26) and later by the 1000 Genomes Project (1), when such functional reference datasets will become available to the field, the interpretation, follow-up and clinical relevance of GWAS results will take a big leap forward. The understanding of the direct consequences of genetic variants to human biology will result in better and more comprehensive interpretation of personal genomes, a step that is likely to transform biomedical research and medical practice.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Louis-Jeantet Foundation, the Swiss National Science Foundation (to E.T.D.) and the European Molecular Biology Organization (EMBO ALTF-337-2010 to H.K.).

REFERENCES

- 1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
- Hindorf, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
- Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
- Dermitzakis, E.T. (2012) Cellular genomics for complex traits. *Nat. Rev. Genet.*, **13**, 215–220.
- Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E. and Lumey, L.H. (2008) Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl Acad. Sci. USA*, **105**, 17046–17049.
- Zhou, V.W., Goren, A. and Bernstein, B.E. (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.*, **12**, 7–18.
- Olovnikov, I., Aravin, A.A. and Fejes Toth, K. (2012) Small RNA in the nucleus: the RNA-chromatin ping-pong. *Curr. Opin. Genet. Dev.*, **22**, 164–171.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, **148**, 458–472.
- Ernst, J., Kheradpour, P., Mikkelson, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.
- Filion, G.J., van Bemmelen, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, **143**, 212–224.
- Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
- McDaniell, R., Lee, B.K., Song, L., Liu, Z., Boyle, A.P., Erdos, M.R., Scott, L.J., Morken, M.A., Kucera, K.S., Battenhouse, A. *et al.* (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. *Science*, **328**, 235–239.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.
- Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H. and Nadeau, J.H. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.*, **11**, 446–450.
- Jirtle, R.L. and Skinner, M.K. (2007) Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.*, **8**, 253–262.
- Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E. *et al.* (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, **482**, 390–394.
- Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y. and Pritchard, J.K. (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.*, **12**, R10.
- Ng, S.F., Lin, R.C., Laybutt, D.R., Barres, R., Owens, J.A. and Morris, M.J. (2010) Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature*, **467**, 963–966.
- Rakyan, V.K., Down, T.A., Balding, D.J. and Beck, S. (2011) Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.*, **12**, 529–541.
- Rakyan, V.K., Beyan, H., Down, T.A., Hawa, M.I., Maslau, S., Aden, D., Daunay, A., Busato, F., Mein, C.A., Manfras, B. *et al.* (2011) Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis. *PLoS Genet.*, **7**, e1002300.

24. Breitling, L.P., Yang, R., Korn, B., Burwinkel, B. and Brenner, H. (2011) Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.*, **88**, 450–457.
25. Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A. *et al.* (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.*, **8**, e1002629.
26. International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.