Genome **Biology**

RESEARCH HIGHLIGHT

DNA methylation, genotype and gene expression: who is driving and who is along for the ride?

Meaghan J Jones[1], Anthony P Fejes[1] and Michael S Kobor[1,2]*

Abstract

A new study integrates genome-wide SNP genotyping, RNA-Seq and DNA methylation in human cells, revealing their relationships and posing new questions about causality.

Introduction

Over the past decade, massive amounts of genome-wide data have been generated for gene expression, genotyping and, more recently, DNA methylation. Each of these three display inter-individual variation and thus are candidates for the discovery of variants correlated with quantitative traits. They are therefore currently being explored for associations with diseases, environmental exposures, specific drug interventions, and many other phenotypic effects.

One of the major challenges in these rapidly evolving -omics fields is how to integrate datasets in a way that will reveal how the three types of variation inter-relate. As we uncover more details about this network of interactions, we are beginning to discover how disruption of gene expression, DNA methylation or genotype can affect one another on a genome-wide scale. Currently, questions remain as to what extent all three types of variability are inter-dependent, how these relationships vary between cell types and individuals, and how these relationships change throughout an individual's lifetime.

Linking SNP genotype, DNA methylation and gene expression

A recent article by Dermitzakis and colleagues has taken a step forward in the integration of these disparate types of data by combining data from RNA-seq, SNP genotyping and the Illumina 450K Human DNA Methylation platform to investigate the relationship between gene expression, genotype and CpG methylation [1]. Using samples of three different cell types isolated from the cord blood of 195 newborn infants, they examined concordance between the three types of data to find areas of association. Expression quantitative trait loci (eQTLs) are correlations between SNPs and gene expression, methylation quantitative trait loci (mQTLs) are correlations between SNPs and methylation, and expression quantitative trait methylations (eQTMs) are correlations between gene expression and methylation. The numbers of eQTLs, mQTLs and eQTMs differed between the three tissue types, with mQTLs accounting for the most sites and eQTMs the least [1]. Consistent with previous studies, these associations account for only a small fraction of the assayed CpG sites, SNPs and expressed genes.

In keeping with previous reports, the authors found that within a single sample, increased methylation at promoters is associated with decreased gene expression across all genes [1-3]. As was previously published elsewhere, when comparing specific sites across multiple samples, some show the opposite trend: increased gene expression with increased methylation [2]. The authors of this article went one step further, subdividing their sites based on whether they showed a positive or negative correlation between promoter DNA methylation and gene expression across individuals. They found that, regardless of which direction the across-individual correlation was, the across-gene correlation within each individual was always negative. That is, within an individual, analysis of even the positive eQTMs revealed a negative correlation with gene expression.

The authors then showed that the CpG sites that were positively correlated with expression across individuals were significantly less likely than their negatively correlated counterparts to be found in CTCF binding sites, enhancers and promoters, particularly non-CpG island promoters [1]. This implies that positive eQTM CpGs are using other mechanisms to influence expression levels.

Another interesting finding in this study involved modeling the causative relations of all three types of data. The authors used SNPs as the starting point, since SNPs are the least likely to change over time, and tested

*Correspondence: msk@cmmt.ubc.ca
[1]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, BC, V5Z 4H4, Canada
Full list of author information is available at the end of the article

whether a SNP is more likely to: (1) affect a methyl site that, in turn, affects gene expression levels; (2) affect gene expression levels that, in turn, affects DNA methylation; or (3) affect both gene expression and methylation independently of one another. Surprisingly, in all three tissues, it is more likely that a SNP affects methylation and expression independently, with SNP to expression to methylation being the least likely in all cases [1]. To determine the underlying mechanism by which SNPs affect both of these in an independent manner, they examined transcription factor (TF) expression levels across individuals, assaying whether an increased or decreased presence of TFs affects both methylation and expression. They examined eQTMs that overlapped with known TF binding sites, and found an enrichment of significant associations with TF expression levels [1]. Thus, if a SNP interferes with or alters a TF binding site, it could potentially affect both DNA methylation and gene expression independently.

The authors also showed that CpG sites with the greatest methylation differences between tissues were enriched for both mQTLs and eQTMs. This indicates that tissue-variable sites contribute to inter-individual variation as well. It follows that some of these eQTMs are a result of gene expression affecting local DNA methylation, and thus differences in the transcriptome between cell types results in different eQTMs. For mQTLs, though, the relationship is less clear, as this suggests SNP-dependent methylation at a given site is also tissue-dependent.

Integrating development, aging and the environment

The analysis done in this study yields important insights into how DNA methylation interacts with both genomic variants and gene expression. However, relatively little of the variation in the dataset could be explained by simple models accounting for these three players. In the context of epigenetics, it is also necessary to consider the changes observed during an individual's development, and the influence of the environment upon these changes.

It is tempting to speculate on how or whether the interactions between gene expression, DNA methylation and genotype vary over an individual's lifetime. It has recently been shown that there is massive shifting of patterns of DNA methylation and histone modifications in the brain during restructuring and learning [4,5]. It follows that the interactions of methylation with genotype and gene expression might also change during that time for those CpGs that are linked to allelic variation and gene expression. Given the models presented, it is possible, for example, that changes in gene expression with age result in the changes in DNA methylation patterns that we associate with aging. Alternatively, it

could also be that aging directly remodels DNA methylation, resulting in changes in gene expression.

It has been clearly shown not only that DNA methylation changes over time, but also that genetic variants can affect the rate at which a person's methylome changes with age [5,6]. It is entirely possible that understanding the mechanisms by which DNA methylation, gene expression and genomic variation interact over the course of a person's life could yield valuable clues to improving human health and opening new avenues of disease prevention. A similar analysis to that performed in the study by Dermitzakis and colleagues expanded to cover a wide range of ages has the potential to unlock many of the mysteries associated with age-related cellular changes.

It is also possible that the environment can be integrated into the proposed models as a global modifier of all three sources of variation, given that exposures to environmental factors can alter gene expression levels as well as DNA methylation [7]. One of the questions asked when assaying a correlation between an environmental or social exposure and an epigenetic pattern is whether the pattern is a result of the exposure or whether it is a sign that the exposure has affected something else in the genome. The current work shows that patterns of DNA methylation that are associated with gene expression levels, for example, are more frequently causative of the gene expression change as opposed to being a result of the change for all three tissues. This is an exciting observation for those who study developmental- and environmental-related changes in DNA methylation, as it describes a potential mechanism by which experiences and exposures can influence the epigenome. At the same time, it is important to note that a large proportion of the eQTMs examined were passive DNA methylation patterns established as a result of gene expression changes.

Implications for population epigenetics

The incorporation of numerous complementary methods together in this study is an important milestone in the maturity of the field of epigenetics, signaling that we are able to unite disparate information sources together to identify patterns that would be invisible when investigated through the lens of a single method. Detailed mapping of these passive versus active DNA methylation patterns in different tissues will be an important next step. It will be vital to identify what determines whether methylation is causative of gene expression changes versus merely being an indicator of transcriptional status. In current population epigenetics studies, large numbers of putative gene associations are being identified for such disparate areas as disease, stress, social history and mental health. If we do not know which of the associations are causative of gene expression changes and

which are merely passengers, it will be difficult to differentiate between potential biomarkers and druggable gene targets.

Another important implication for the field of population epigenetics that comes from this study is that the association between DNA methylation and gene expression levels is not likely to be allele-specific. Thus, 50% methylation and reduced gene expression at a specific site are much more likely to mean that half of the cells collected were methylated and show lower expression than that all the cells are methylated only on a single chromosome with allele-specific expression.

It is curious that, in the article under discussion, no mention is made of how many eQTLs, eQTMs and mQTLs overlap between the tissues or even whether specific CpGs are found on the eQTM and mQTL lists. In previous studies examining more than one tissue, some eQTMs did overlap between tissues [8]. It would be very interesting to note how many of each are found in one or more tissues. It could be predicted, for example, that eQTMs found in all three tissues might have a broader function and could be more stable during development. These cross-tissue eQTMs may be stronger targets for mechanistic experiments to determine through which pathways gene expression and DNA methylation are being coordinated. In addition, mQTLs have been shown to vary across ethnic groups, suggesting the possibility of multi-gene interactions controlling polymorphic DNA methylation [8].

Given the interesting data presented in the article, it will be very exciting to apply a similar approach with genome-wide data for histone modifications, and add this layer to the models proposed. It is likely that histone modifications are at the interface between some of the proposed interactions, mediating the relationship between DNA methylation and gene expression, for example; adding this layer to the current data will therefore help to further clarify these relationships.

In the context of human disease, the current work presents the first evidence for interactions between genetic, epigenetic and transcriptional variants on a genome-wide scale. Studies such as this, integrating the current wealth of genome-wide technologies, are poised to bring us further information on how and whether multiple types of variation are collaborating to modify disease and disease risk. There are currently few published examples of specific variants interacting to affect disease, but one example showed DNA methylation and genotype interacting as partners to affect the risk of

rheumatoid arthritis [9]. It will be interesting to see whether some of the unaccounted-for variation in well-studied diseases may be discovered to be an example of these molecular interactions between genotype, gene expression and DNA methylation.

Abbreviations
eQTLs, expression quantitative trait loci; eQTM, expression quantitative trait methylation; mQTLs, methylation quantitative trait loci; SNP, single nucleotide polymorphism; TF, transcription factor.

Competing interests
The authors declare that they have no competing interests

Author details
[1]Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, Department of Medical Genetics, University of British Columbia, BC, V5Z 4H4, Canada. [2]Human Early Learning Partnership, School of Population and Public Health, University of British Columbia, BC, V6T 1Z3, Canada.

References
1. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, Bryois J, Giger T, Romano L, Planchon A, Falconnet E, Bielser D, Gagnebin M, Padioleau I, Borel C, Letourneau A, Makrythanasis P, Guipponi M, Gehrig C, Antonarakis SE, Dermitzakis ET: **Passive and active DNA methylation and the interplay with genetic variation in gene regulation.** *eLife* 2013, **2**:e00523.
2. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS: **Factors underlying variable DNA methylation in a human community cohort.** *Proc Natl Acad Sci U S A* 2012, **109 Suppl 2**:17253-17260.
3. Han H, Cortez CC, Yang X, Nichols PW, Jones PA, Liang G: **DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter.** *Hum Mol Genet* 2011, **20**:4299-4310.
4. Shulha HP, Cheung I, Guo Y, Akbarian S, Weng Z: **Coordinated cell type-specific epigenetic remodeling in prefrontal cortex begins before birth and continues into early adulthood.** *PLoS Genet* 2013, **9**:e1003433.
5. Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, Behrens MM, Ecker JR: **Global epigenomic reconfiguration during mammalian brain development.** *Science* 2013. doi:10.1126/science.1237905.
6. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan J-B, Gao Y, Deconde R, Chen M, Rajapakse I, Friend S, Ideker T, Zhang K: **Genome-wide methylation profiles reveal quantitative views of human aging rates.** *Mol Cell* 2013, **49**:359-367.
7. Meaney MJ: **Epigenetics and the biological definition of gene x environment interactions.** *Child Dev* 2010, **81**:41-79.
8. Fraser HB, Lam LL, Neumann SM, Kobor MS: **Population-specificity of human DNA methylation.** *Genome Biol* 2012, **13**:R8.
9. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekström TJ, Feinberg AP: **Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis.** *Nat Biotechnol* 2013, **31**:142-147.