Incorporating alternative splicing and mRNA editing into the genetic analysis of complex traits

*Musa A. Hassan and Jeroen P. J. Saeij**

The nomination of candidate genes underlying complex traits is often focused on genetic variations that alter mRNA abundance or result in non-conservative changes in amino acids. Although inconspicuous in complex trait analysis, genetic variants that affect splicing or RNA editing can also generate proteomic diversity and impact genetic traits. Indeed, it is known that splicing and RNA editing modulate several traits in humans and model organisms. Using high-throughput RNA sequencing (RNA-seq) analysis, it is now possible to integrate the genetics of transcript abundance, alternative splicing (AS) and editing with the analysis of complex traits. We recently demonstrated that both AS and mRNA editing are modulated by genetic and environmental factors, and potentially engender phenotypic diversity in a genetically segregating mouse population. Therefore, the analysis of splicing and RNA editing can expand not only the regulatory landscape of transcriptome and proteome complexity, but also the repertoire of candidate genes for complex traits.

Keywords:
alternative splicing; quantitative genetics; RNA editing

DOI 10.1002/bies.201400079

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA

***Corresponding author:**
Jeroen P. J. Saeij
E-mail: jsaeij@mit.edu

Abbreviations:
AS, alternative splicing; **eQTL,** expression quantitative trait locus; **QTL,** quantitative trait locus; **sQTL,** splicing quantitative trait locus.

Introduction

Most genetic traits in humans and model organisms are modulated by polymorphisms (nucleotide and structural variants) at multiple genes (polygenic) and do not follow simple Mendelian patterns of inheritance. Even where a polymorphism at a single gene modulates a genetic trait (monogenic), the genetic basis for the trait can be convoluted by the effect of modifier genes. These complex genetic networks can further be compounded by interactions with environmental and behavioral factors. Consequently, the identification of causal genes for most genetic traits is complicated. The conventional approach to finding causal genes is to use positional cloning to identify genomic regions that correlate with the trait under study (quantitative trait locus, QTL) [1] and examine all the genes under the QTL for putative candidates. Principally, this entails the identification of genes that contain polymorphisms that either change protein sequence and function (non-conservative non-synonymous polymorphisms) or cause variable transcriptional profiles between individuals that display the different phenotypes [2, 3]. However, a single QTL, even after fine mapping, often contains multiple genes, thus making the identification of the causative gene an arduous task. Even where a causative gene has been identified, the biological mechanisms that link genotype and phenotype often take a long time to elucidate.

The observation that, between genetically divergent individuals, cellular mRNA levels are variable and can be linked to specific genomic regions (genetical genomics) [4] has revolutionized complex trait genetics and provided an option for detecting all variants affecting gene expression regardless of functional annotation. The established approach in genetical genomics is to treat mRNA abundance as a quantitative trait, and then, using linkage analysis, determine the genomic regions that regulate the expression of each transcript (expression quantitative trait locus, eQTL). The eQTL can then be positionally characterized as cis or trans based on its genomic location relative to the relevant gene; cis-eQTL are proximal to the gene they regulate (often ≤10 Mb) while trans-eQTL are located further away from the relevant gene (often >10 Mb or on a different chromosome from the gene it regulates). The eQTL

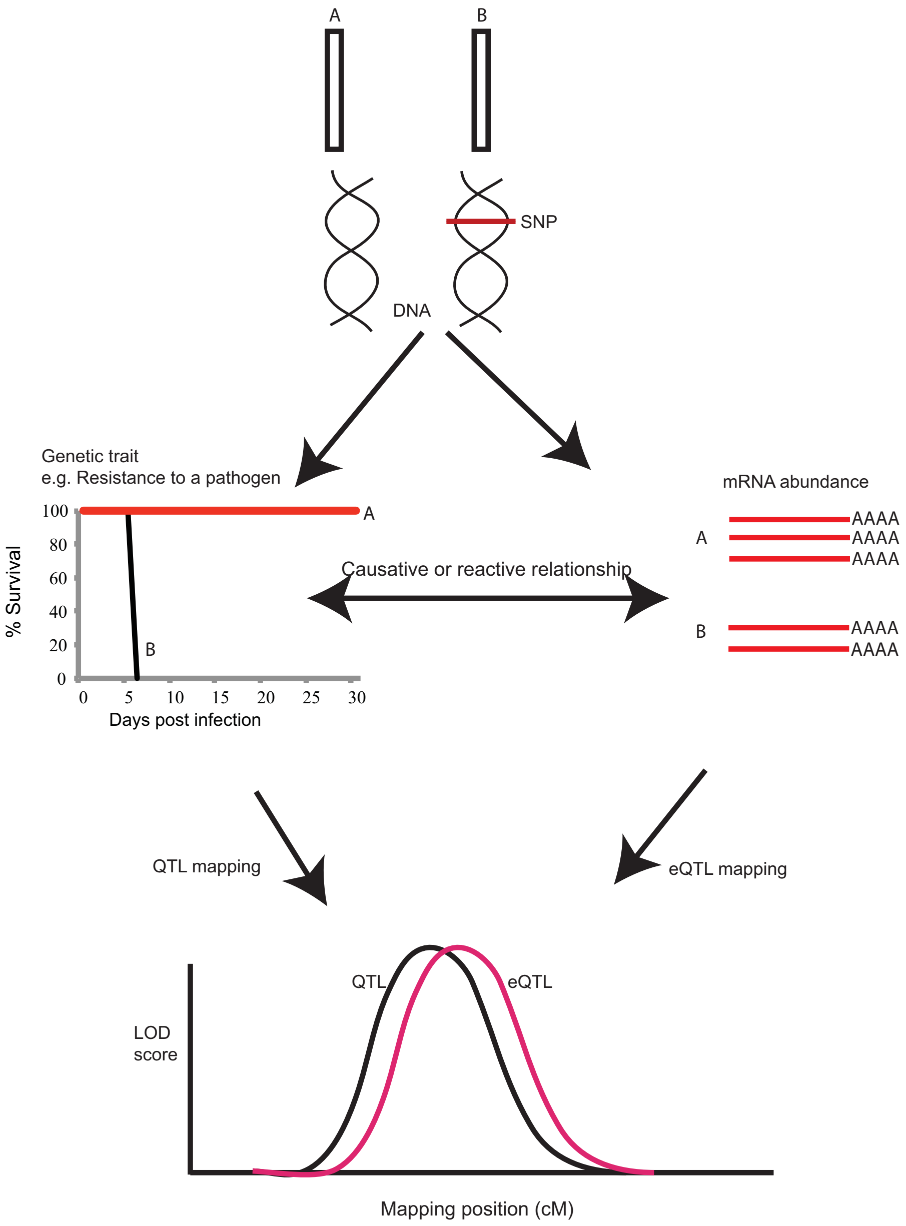

Figure 1. Genetical genomics. Shown are two individuals "A" and "B" segregating at a single locus (single nucleotide polymorphism, SNP). The SNP affects mRNA abundance, which in turn modulates a genetic trait, e.g. survival time after exposure to a pathogen. In linkage analysis, the expression quantitative trait locus (eQTL) (red curve) responsible for gene expression and the quantitative trait locus (QTL) (black curve) responsible for the genetic trait should co-localize at the SNP. By using functional and network analysis of genes physically located in the QTL region and the eQTL, one can identify the causative gene and molecular mechanism underlying the genetic trait. However, even in the absence of a causal relationship between mRNA abundance and the genetic trait, the eQTL and QTL would still co-localize at the SNP.

can then be integrated into the genetic analysis of complex traits to find putative candidate genes. In this last step, the general assumption is that, if a genetic variant is regulating the expression of a gene(s), which in turn modulates a complex trait, then the eQTL and the complex trait QTL should co-localize at the causal locus [5, 6] (Fig. 1). However, occasionally there is deviation from this causal relationship between gene expression and a complex trait, which can convolute the identification of candidate genes even with the integration of genetical genomics. For instance, a genetic variant might independently modulate gene expression and a complex trait (independent relationship) or a genetic variant can modulate a complex trait that may in turn result in changes in gene expression (reactive relationship in which case the gene expression is an effect of, rather than the driver of, the complex trait). In the causal, independent, and reactive relationships [6], the eQTL and QTL would still co-localize at the variant. Nevertheless, several bioinformatic and network analysis approaches have been used successfully to deconstruct this complex relationship and logically identify candidate genes [5–8].

A major drawback to genetical genomics is that it presupposes that mRNA abundance represents the entirety of transcriptome complexity. Emerging empirical evidence indicates that several factors, such as alternative splicing (AS) and mRNA editing, not only contribute to transcriptome diversity but also are variable among individuals [9]. In fact, often protein abundance does not mirror the steady state transcript levels [10, 11]. Consequently, we submit that in addition to transcript abundance, protein abundance, AS, and mRNA editing can be integrated with genetical genomics to identify putative candidate genes that modulate genetic traits. Below, we provide a brief review on the putative influence of AS and DNA/RNA editing on quantitative trait genetics.

Alternative splicing modulates a variety of cellular phenotypes

In eukaryotes, most protein coding genes are transcribed as precursor messenger RNA (pre-mRNA); in which the protein coding regions (exons) are interrupted by non-coding regions (introns) [12]. Consequently, the pre-mRNA must be processed into mature mRNA (mRNA) before translation into a protein. A key step in the processing of pre-mRNA is splicing, which entails the removal of intervening introns to assemble the protein-coding exons into mRNA [13]. Splicing, catalyzed by a protein mega-complex – spliceosome – that is sequentially assembled at specific sequences on exon-intron junctions (splice junction) [13], is a dynamic and highly regulated process [14]. Often, a pre-mRNA contains more than one putative splice site, any of which can be a docking site for the spliceosome complex. The choice of a splice junction is determined by, amongst others, the "mooring" sequences around the splice site that are recognized by the spliceosome complex [15, 16]. Generally, in certain tissues or cell types, a single splice site is preferentially used in processing the pre-mRNA, resulting in the predominance of a single mRNA isoform in the transcriptome, the constitutive isoform. However, sometimes – for a variety of reasons such as altered cellular physiology and cell type – alternative splice sites may preferentially produce the mRNA, resulting in the abundance of alternative transcript isoforms. AS, the generation of multiple transcript isoforms from a single gene, is pervasive in eukaryotes, and impacts several aspects of eukaryotic biology, including responses to environmental and pathogen exposure [17, 18]. For example it has been reported that mitotic arrest, induced by drugs or siRNA, is often characterized by the AS of various pro-apoptotic transcripts [19]. A recent study indicates that some tissue-specific AS signatures are conserved across species [20].

There are suggestions that differential mRNA splicing may be more important than differential gene expression in modulating human genetic traits [21]. A case in point is the variable efficacy of statins that is modulated by differences in 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR) isoform usage [22]. Even though the examples highlighted above accentuate a putative role for AS in quantitative trait genetics, large-scale studies investigating differential isoform usage in genetically divergent individuals or studies that use the variability in AS to identify candidate genes modulating genetic traits are still rare [23]. This paucity in large-scale isoform usage analysis can partly be attributed to the lack of reliable means for measuring differential AS without the confounding effect of the fluctuations in the expression of the parent transcripts. Previously, most studies relied on exon arrays to capture the inclusion or exclusion of exons in mRNA pools [24–26]. However, in addition to the common limitations of array-based gene expression technology, this method cannot reliably distinguish between differential AS and overall transcript abundance. Additionally, while exon-arrays may capture exon inclusion or exclusion, there are several types of AS, such as alternative splice acceptor or donor sites, that do not change exon counts, and thus cannot be accurately captured by exon-arrays. A surge in bioinformatics tools that can, with improved certitude, estimate differential isoform usage [27, 28] is currently fueling studies that incorporate AS in complex trait studies.

Genetic analysis of isoform usage can facilitate the identification of candidate genes for complex traits

Although non-synonymous polymorphisms dominate complex trait genetic analysis, polymorphisms in the splice donor/acceptor sites, intronic splice enhancers/suppressors (ISE/ISS), or branch point can lead to AS of pre-mRNA, leading to alternative transcript and protein isoforms. However, the polymorphisms that influence pre-mRNA splicing may be located in introns, and are often considered inconsequential in modulating transcript and protein abundance or function. This may lead to the erroneous categorization of such polymorphisms as non-causal. Take as an example gene "A" that does not have any non-synonymous polymorphisms, and which produces the constitutive and alternative isoform A1 and A2, respectively, one of which is non-functional. Consider two models for gene "A" isoform usage in two individuals exhibiting different phenotypes for a genetic trait. In the first model, the usage of A1 and A2 is the same, thus "A" is non-variable. In the second model, due to a polymorphism that affects splicing of "A", one individual predominantly expresses isoform A1 while the second individual mostly expresses isoform A2, without changing the overall transcript abundance of "A" (Fig. 2). In both models, unless we examine isoform usage, it is impossible to conceptualize the contribution of gene "A" to the phenotypic difference. In the first model, based on the absence of non-synonymous polymorphisms, it is unlikely that the trait under investigation is influenced by gene "A". In the second model, with everything else being constant, it is possible that gene "A" is a putative candidate, even though it lacks a non-synonymous polymorphism.

Information about alternative isoforms for individual transcripts modulating

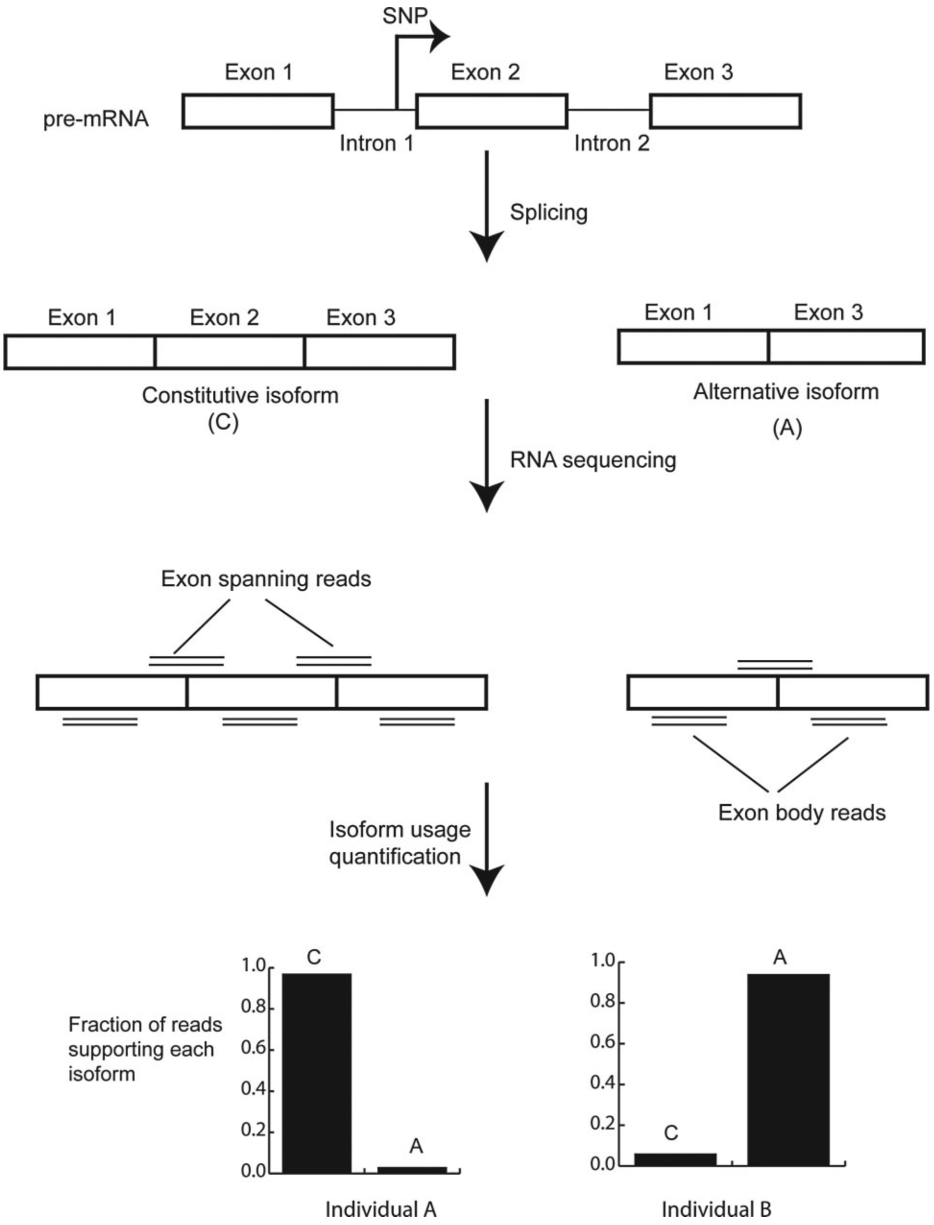

Figure 2. Genetics of alternative splicing. During the processing of pre-mRNA, multiple splice-junctions have the potential to be bound by the spliceosome. However, under normal physiology, a single splice junction is preferred, resulting in the dominance of the constitutive (C) over the alternative (A) transcript isoform in the transcriptome. A polymorphism proximal to the splice site may affect the binding of the spliceosome to the canonical splice site, leading to variable isoform usage between individuals. By quantifying the reads that map to the exon-exon junctions and the exon-bodies, one can estimate the fraction of reads that support each isoform in genetically divergent individuals. These isoform usage estimates can then be used in linkage analysis, similarly to mRNA abundance, to identify the genomic regions modulating splicing.

differences in complex traits is now commonplace [23, 29–31]. However, for AS to be incorporated into quantitative trait genetics, it must be variable, heritable, and feasible to reliably capture the genome-wide splicing architecture. In fact, it can be argued that the success of genetical genomics and its incorporation into the genetics of complex traits is largely due to the heritability of mRNA abundance, not to mention the simplicity of performing large-scale transcriptional analysis. Recently, we integrated isoform usage and linkage analyses in a genetically divergent set of recombinant inbred (RI) mice, and showed that isoform usage varies with both the macrophage genetic background and physiological state [9]. For instance, we observed that the C-type lectin domain family 7 member a (*Clec7a* or Dectin-1) gene, from which an alternative transcript that lacks the 4th exon can be produced [32], is differentially spliced in bone marrow-derived macrophages (BMDM) obtained from the classical laboratory inbred A/J (AJ) and C57BL/6J (B6) mice. Even though the overall *Clec7a* expression is non-variable and there are no known non-synonymous polymorphisms between AJ and B6, due to a cis-acting polymorphism, we observed that B6 macrophages expressed mostly the truncated alternative *Clec7a* isoform. Consequently, despite the lack of non-synonymous polymorphisms, *Clec7a* – which has been implicated in the immune response to a variety of pathogens including *Salmonella* and *Candida albicans* – is a viable candidate gene for the differential response of AJ and B6 to such pathogens. Indeed, the truncated *Clec7a* isoform has previously been associated with increased B6 susceptibility to *Coccidioides* [32]. Based on the genomic location of the genetic variant and the splicing event, we further categorized the genetic loci that modulate splicing (splicing QTL, sQTL) as cis or trans. cis-sQTL, like the *Clec7a* sQTL, are generally defined as proximal (in our case $\leq$10 Mb) to the alternatively spliced exon, while trans-sQTL were located ($>$10 Mb) away from the relevant splicing event. The 5 RNAs and about 300 distinct proteins that constitute the splicing complex [33] can affect splicing in trans. In eQTL analysis, when a polymorphic transcription factor regulates the expression of multiple genes in trans, a trans-eQTL hotspot is observed at the physical location of the transcription factor [34]. Similarly, when polymorphisms in these trans-splicing factors affect the splicing of multiple transcripts, a trans-sQTL hotspot can be observed at the physical position of these trans-factors [9]. If these trans-sQTL co-localize with a complex trait QTL, the local splice factor can be considered a candidate gene for the trait. In addition, similar to transcript analysis, sQTL in the trans-sQTL hotspot can be used in network analysis, in which the splice factor and the spliced genes form the network core and nodes, respectively, to delineate the molecular mechanisms modulating a trait. A similar AS linkage analysis in humans recently identified several causative single nucleotide polymorphisms (SNPs), 13 of which were associated with about 84 common human genetic traits [23], underscoring the significant role that AS may play in determining phenotypic diversity in humans.

In addition to the quantitative trait genetics value of AS, differential splicing can also be useful in delineating the cellular response to various stimuli. For example using the same BMDM from AJ and B6, we observed that compared to interferon gamma (IFNG)-stimulated BMDM, *Toxoplasma gondii*-infected murine macrophages expressed mostly the unstable alternative isoform of sterile alpha motif domain- and HD domain-containing gene (*Samhd1*), which lacks the 14th exon and produces a catalytically inactive protein [35]. However, the expression level and isoform usage for *Samhd1* was not variable between the mouse strains following individual stimulations. SAMHD1 is a triphosphohydrolase that depletes the cellular pool of deoxynucleoside triphosphates, and has been implicated in cancer pathogenesis and intracellular retroviral replication [36, 37]. *T. gondii* is an obligate intracellular parasite that infects virtually all nucleated cells, from which it needs to scavenge purines because it is a purine auxotroph. It is thus plausible that by promoting the canonical splicing of *Samhd1*, IFNG, which is indispensable in the resistance to *T. gondii*, controls intracellular parasite replication by depleting cellular nucleotides, a possibility that is otherwise indiscernible when we only consider the overall expression of *Samhd1*. The differential splicing of *Samhd1* may also modulate intracellular retroviral replication, which also relies on the availability of intracellular pools of nucleotides. Indeed, several studies have indicated that *Samhd1* is a key regulator for HIV-1 replication in murine cells [38, 39]; cells lacking *Samhd1* are more permissible to the virus [38]. These examples highlight the possibility of expanding the putative candidate gene pools in complex trait genetic analysis. However, because AS tends to be tissue and/or cell-specific, one needs to be cautious when establishing links between splicing events and phenotypes, particularly where there is divergence on tissue and/or cell types. Indeed, this is one factor that may continue to hinder the incorporation of AS in quantitative trait genetics, particularly in humans, where it can often be difficult to obtain different types of tissues and/or cells.

DNA and/or mRNA editing can influence individual phenotypic differences

Besides AS, eukaryotic transcriptome diversity can further be achieved through DNA and/or RNA editing. Editing, which involves the interchange of two-ring purines or one-ring pyrimidines in RNA post-transcriptionally or in DNA, is generally catalyzed by either adenosine deaminase acting on RNA (ADAR) or apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC). While the former deaminates adenosine to inosine, which is interpreted as guanosine by the mRNA translation machinery (A-to-G), the latter deaminates cytosine to thymidine in DNA or uracil in RNA (C-to-T/U) [40–42]. In addition to APOBECs, cytidine deaminases also include the activation-induced cytidine deaminase (AID or AICDA) that primarily deaminates cytidine to thymidine in DNA (C-to-T). The APOBEC deaminases include APOBEC1, APOBEC2, APOBEC3, and APOBEC4. Due to AS, APOBEC3 exists in various alternative isoforms in humans (A3A, A3B, A3C, A3D, A3F, A3G, and A3H). While APOBEC1 and 3 have been associated with several phenotypes, the biological function for APOBEC2 and 4 is yet to be defined. Nevertheless, ADAR-catalyzed editing dominates in higher eukaryotes [43, 44]. Although *Adar* is the best characterized, the mammalian genome encodes two additional *Adar* genes (*Adarb1* and *Adarb2*) [45, 46]. Editing has been implicated in a variety of phenotypes, including antibody diversification [47], cell proliferation [41], and responses to viruses [48]. Indeed, editing-induced hypermutation of the viral genome has been implicated in the failure of the HIV provirus to exit latency in the memory pool of CD4^{+} T cells [49]. Considering that viral latency significantly affects HIV therapy [50, 51], it can be presumed that editing modulates the efficacy of antiretroviral drugs. Besides the editing of the viral genome, DNA/RNA editing can also influence the cellular response to pathogens by directly targeting and influencing the function of host genes. However, apart from the well-studied ADAR enzyme, the biological significance of APOBEC-catalyzed editing on host genes is not well understood. APOBEC1-mediated editing has been noted to preferentially occur at the mRNA 3′UTR [9, 42], within the miRNA-binding seed sequence [42]. It is thus plausible that editing influences mRNA stability, translation, cellular localization, or 3′UTR splicing, which can have profound effects on the host biological pathways and response to pathogens. Importantly, a recent study observed that some viruses, such as human cytomegalovirus, produce miRNAs that specifically bind and degrade host mRNAs [52]. Indeed, the literature is replete with repertoires of viral miRNAs that target host genes within important biological pathways such as cell growth and differentiation [53, 54]. Although anecdotal, together with the preferential editing within the miRNA mooring sequence at the mRNA 3′UTR and the correlation of editing with viral replication, it is plausible that editing evolved as a mechanism to counter viral encoded miRNAs that target host genes.

Polymorphisms in the editosome-binding sequences or the deaminases are likely to result in differential level of DNA/RNA editing events between individuals, and can be used as genetic markers in the analysis of quantitative genetic traits. However, since several factors, including the RNA duplex structure [55, 56] converge to regulate ADAR-mediated editing, it maybe difficult to incorporate A-to-I editing activity into quantitative trait genetics. Nevertheless, as described above for AS, the main challenge in incorporating editing in quantitative trait genetics is the dearth of evidence that it is either variable or genetic. Even though several studies have explored the DNA or RNA editing background of cells and tissues from a variety of species and cellular physiology [40, 42, 57, 58], few have performed comparative studies on the level of editing in genetically divergent individuals. Recently, we showed that the extent of RNA edited species is variable and genetically transmissible in murine BMDM [9]. This observation raises a few possibilities, including the use of editing as a variable component in the genetic analysis of complex traits. For instance, if a QTL for a genetic trait, such as disease susceptibility or transcript abundance, overlaps a polymorphic editing enzyme, then editing can be included as a possible biological mechanism regulating the trait under study, and the editing enzyme can be considered a viable candidate gene. The former scenario is exemplified by APOBEC1, which is known to modulate plasma cholesterol levels by editing apolipoprotein B (*Apob*) [59]. In our study, we found editing QTL (edQTL) for several editing events, mainly C-to-T, that mapped in trans at the *Apobec1* locus on mouse chromosome 6. Interestingly, this region on chromosome 6 includes QTL for several complex traits such as atherosclerosis, cytomegalovirus resistance, and obesity, all of which are thought to be at least modulated by DNA/mRNA editing. Among the genes edited by APOBEC1, and which contain a trans-edQTL at this locus, is the amyloid precursor protein (*App*), which has been associated with several phenotypes including Alzheimer’s disease and diabetes [60, 61]. We reasoned that a polymorphism in *Apobec1* was responsible for the trans-edQTL hotspot at the *Apobec1* locus. Indeed we found *Apobec1* to be differentially expressed and alternatively spliced between AJ and B6 BMDM, in which both the differential expression and AS map in cis (cis-eQTL and cis-sQTL). We speculated that this differential isoform usage was due to a polymorphism in its splice-junction resulting in an alternative splice acceptor site in the 3rd exon. Although this alternative isoform does not result in an alternative protein isoform, the change in the exon structure can affect *Apobec1* mRNA translation efficiency. It is also possible that the differential expression of *Apobec1* is the source of the variable C-to-T transitions between AJ and B6. Indeed, variable expression of *Apobec* has been associated with differential level of editing in human cells [41]. It is thus plausible that *Apobec1* is the causative gene for traits such as Alzheimer’s disease, which have QTL at the *Apobec1* locus on chromosome 6, and that edited genes such as *App* act downstream of *Apobec1*. Therefore, by comparing the level of edited events in individuals segregating for a phenotype, new candidate genes can be identified.

How can we test these hypotheses?

It is now clear that AS and editing can be used as markers in the genetic analysis of complex traits. However, before splicing or editing factors and alternatively spliced or edited genes are considered candidates, confirmatory experiments must be performed. Since the polymorphisms modulating these events are mostly in non-coding regions or, in the case of editing, involve single nucleotide substitutions; the key challenge is how to perform confirmatory experiments. Linkage analysis will reveal loci that modulate splicing or editing, and when integrated with the complex trait (e.g. susceptibility to disease) QTL, it may reveal loci that regulate splicing or editing and the complex trait under study. With the advent of genome editing technology such as clustered regularly interspersed short palindromic repeats-associated genes (CRISPR/*Cas*), it is now possible to engineer single nucleotide substitutions and test their effects in vitro and in vivo [62, 63]. Take as an example the AS of the *Apobec1* gene, which we speculate is due to a SNP near a splice junction. It is plausible that a guanosine proximal to the splice junction promotes the expression of the truncated *Apobec1* isoform in the B6 BMDM, which in turn leads to lower rates of editing. Therefore, it is possible to investigate isoform usage and editing activity of *Apobec1* by using CRISPR/Cas system to change the guanosine to thymidine in the B6 BMDM. This would essentially introduce a site-specific substitution in the B6 *Apobec1* allele and convert it to an AJ allele without changing the rest of the nucleotide sequence or knocking out the entire gene. We can then investigate quantitative traits, such as susceptibility to cytomegalovirus, with QTLs overlapping the *Apobec1* locus. A similar approach can be used to test the possibility that editing is an evolutionary mechanism directed against viral miRNA-mediated degradation of host genes.

Conclusions and outlook

Naturally occurring genetic polymorphisms may lead to variable and heritable phenotypes within species. In addition to the well-studied fluctuations in transcript levels, such variants may also affect AS and RNA editing, both of which contribute to phenotypic diversity. Variants that modulate splicing or editing are often excluded from quantitative genetic analyses because they are mostly outside of the coding or promoter regions. This omission may hinder the discovery of causative genes, which can partly explain the failure to confirm most candidate variants identified in genome-wide association studies. It can be argued that the complexity of accurately capturing isoform usage and RNA editing has partly impeded the incorporation of these events in quantitative trait genetics. However, we anticipate that the precipitous drop in the cost of DNA/RNA-seq will simplify research into these events and motivate their incorporation into quantitative genetics. Besides, RNA-seq can be utilized for de novo transcript assembly [64, 65] to expand the transcriptional landscape and discover novel coding or non-coding transcripts that impact genetic traits. Even though we have restricted this review to the impact of heritable genetic variants modulating AS and RNA editing, it will be remiss not to mention the important role that epigenetic memory plays in modulating these events (see Box 1). In fact in an ideal experimental setup, in order to capture the complete regulatory landscape underlying a genetic trait, one should capture the transcriptome (transcript levels, AS, and mRNA editing), the proteome, and the epigenetic architecture. Although the cost and sample availability may not permit the capture of all these data in a single experiment, exploiting the increasingly large public data depositories, may help circumvent this problem. However, care must be taken when comparing data obtained from different cells/tissues since these events are often tissue/cell-specific.

Acknowledgments

This work was supported by a NERCE developmental grant (U54-AI057159), National Institutes of Health (R01-AI080621), and by the PEW Scholar in the Biomedical Sciences grants to J.S. M.H. is a Wellcome Trust-Massachusetts Institute of Technology postdoctoral fellow.

References

1. **Kearsey MJ.** 1998. The principles of QTL analysis (a minimal mathematics approach). *J Exp Bot* **49**: 1619–23.
2. **Duerr RH**, **Taylor KD**, **Brant SR**, **Rioux JD**, et al. 2006. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314**: 1461–3.
3. **Yalcin B**, **Willis-Owen SAG**, **Fullerton J**, **Meesaq A**, et al. 2004. Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nat Genet* **36**: 1197–202.

Box 1

The influence of epigenetic memory and population genetics on alternative splicing

There is a growing body of evidence that epigenetics, which include DNA methylation and histone modifications [66], is not only heritable but can also influence quantitative traits [67], such as transcript abundance and splicing [68–70], disease development [71]. For example DNA methylation, which is the addition of a methyl group to cytosine residues, mostly when it is followed by a guanine (CpG) [66], is reported to facilitate exon recognition in cotranscriptional splicing [72]. Because exons are often more highly methylated than introns [72, 73], DNA methylation is often considered a stronger marker for exon-intron boundaries during splicing [74, 75]. Hypothetically, variations in the exon-intron methylation dynamics may result in exon skipping. However, it has been suggested that alternative exon recognition mechanisms may have evolved in genes with equal exon to intron DNA methylation ratios, which is one of the biological explanations given for the variation in alternative splicing patterns between species [20, 76]. Interestingly, it has been shown that DNA methylation is enriched in alternative splice sites and splicing regulatory motifs [77], and that depletion of methylation expedites elongation and splicing of *Hox* transcripts [78], further giving credence to the regulatory potential of DNA methylation in RNA splicing. Similarly, chromatin structure, which involves DNA wrapped around a nucleosome, has a potential to regulate alternative splicing. For instance, nucleosomes, whose positioning along a chromosome can be modified by alterations in chromatin structure [79], are mostly concentrated on exons and are important in the regulation RNA splicing [80]. In addition to marking exons, nucleosomes can also pause transcript elongation, leading to the cotranscriptional recognition of alternative splice sites by the RNA-splicing machinery [80]. Importantly, recent studies have shown that histone modification mechanisms, such as DNA methylation, can be perturbed by naturally occurring genetic variations [81, 82]. Taken together, variations in the genetic architecture are likely to cause differential alternative splicing dynamics between and within species.

Environmental changes can be postulated to modulate alternative splicing at various levels. For instance, there is considerable evidence that changes in the epigenetic state may be induced by environmental cues such as temperature and nutritional changes [83–85]. When these lead to genetic diversification between populations, we can expect alternative splicing to vary too. In fact, environmental factors, such as altitude and temperature, are known to contribute to genetic stratification in populations [86]. Additionally, due to genetic drift and/or natural selection, the allele frequencies of SNPs modulating alternative splicing may vary between populations. This may in turn cause variation in the level of alternative splicing between such populations. Therefore, it would be interesting to review the large genomics dataset compiled from the human genome diversity cell line panel [87, 88] for evidence of RNA splicing variation between populations.

4. **Jansen RC**, **Nap JP.** 2001. Genetical genomics: the added value from segregation. *Trends Genet* **17**: 388–91.
5. **Schadt EE.** 2006. Novel integrative genomics strategies to identify genes for complex traits. *Anim Genet* **37**: 18–23.
6. **Schadt EE**, **Lamb J**, **Yang X**, **Zhu J**, et al. 2005. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* **37**: 710–7.
7. **Segal E**, **Shapira M**, **Regev A**, **Pe'er D**, et al. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34**: 166–76.
8. **Wu C**, **Delano DL**, **Mitro N**, **Su SV**, et al. 2008. Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. *PLoS Genet* **4**: e1000070.
9. **Hassan MA**, **Butty V**, **Jensen KD**, **Saeij JP.** 2014. The genetic basis for individual differences in mRNA splicing and APOBEC1 editing activity in murine macrophages. *Genome Res* **24**: 377–89.
10. **de Sousa Abreu R**, **Penalva LO**, **Marcotte EM**, **Vogel C.** 2009. Global signatures of protein and mRNA expression levels. *Mol Biosyst* **5**: 1512–26.
11. **Vogel C**, **Marcotte EM.** 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–32.
12. **Griffiths AJF**, **Miller JH**, **Suzuki DT**, **Lewontin RC**, et al. 2000. *An Introduction to Genetic Analysis*, 7th edition.
13. **Kornblihtt AR**, **Schor IE**, **Allo M**, **Dujardin G**, et al. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**: 153–65.
14. **Irimia M**, **Blencowe BJ.** 2012. Alternative splicing: decoding an expansive regulatory layer. *Curr Opin Cell Biol* **24**: 323–32.
15. **Anant S**, **Davidson NO.** 2000. An AU-rich sequence element (UUUN[A/U]U) downstream of the edited C in apolipoprotein B mRNA is a high-affinity binding site for Apobec-1: binding of Apobec-1 to this motif in the 3′ untranslated region of c-myc increases mRNA stability. *Mol Cell Biol* **20**: 1982–92.
16. **Bahn JH**, **Lee JH**, **Li G**, **Greer C**, et al. 2012. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res* **22**: 142–50.
17. **Tarn WY.** 2007. Cellular signals modulate alternative splicing. *J Biomed Sci* **14**: 517–22.
18. **Wang ET**, **Sandberg R**, **Luo S**, **Khrebtukova I**, et al. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470–6.
19. **Moore MJ**, **Wang QQ**, **Kennedy CJ**, **Silver PA.** 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142**: 625–36.
20. **Merkin J**, **Russell C**, **Chen P**, **Burge CB.** 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**: 1593–9.
21. **Heinzen EL**, **Yoon W**, **Weale ME**, **Sen A**, et al. 2007. Alternative ion channel splicing in mesial temporal lobe epilepsy and Alzheimer's disease. *Genome Biol* **8**: R32.
22. **Medina MW**, **Krauss RM.** 2009. The role of HMGCR alternative splicing in statin efficacy. *Trends Cardiovasc Med* **19**: 173–7.
23. **Lee Y**, **Gamazon ER**, **Rebman E**, **Lee S**, et al. 2012. Variants affecting exon skipping contribute to complex traits. *PLoS Genet* **8**: e1002998.
24. **Purdom E**, **Simpson KM**, **Robinson MD**, **Conboy JG**, et al. 2008. FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics* **24**: 1707–14.
25. **Subbaram S**, **Kuentzel M**, **Frank D**, **DiPersio CM**, et al. 2010. Determination of alternate splicing events using the Affymetrix exon 1.0 ST arrays. *Methods Mol Biol* **632**: 63–72.
26. **Chen P**, **Lepikhova T**, **Hu YZ**, **Monni O**, et al. 2011. Comprehensive exon array data processing method for quantitative analysis of alternative spliced variants. *Nucleic Acids Res* **39**: e123.
27. **Katz Y**, **Wang ET**, **Airoldi EM**, **Burge CB.** 2011. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**: 1009–15.
28. **Trapnell C**, **Hendrickson DG**, **Sauvageau M**, **Goff L**, et al. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**: 46–53.
29. **Lee RM**, **Hirano K**, **Anant S**, **Baunoch D**, et al. 1998. An alternatively spliced form of apobec-1 messenger RNA is overexpressed in human colon cancer. *Gastroenterology* **115**: 1096–103.
30. **Heinzen EL**, **Ge D**, **Cronin KD**, **Maia JM**, et al. 2008. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol* **6**: e1.
31. **Stambolian D**, **Wojciechowski R**, **Oexle K**, **Pirastu M**, et al. 2013. Meta-analysis of genome-wide association studies in five cohorts reveals common variants in RBFOX1, a regulator of tissue-specific splicing, associated with refractive error. *Hum Mol Genet* **22**: 2754–64.
32. **Jimenez-A MD**, **Viriyakosol S**, **Walls L**, **Datta SK**, et al. 2008. Susceptibility to Coccidioides species in C57BL/6 mice is associated with expression of a truncated splice variant of Dectin-1 (Clec7a). *Genes Immun* **9**: 338–48.
33. **Nilsen TW.** 2003. The spliceosome: the most complex macromolecular machine in the cell? *BioEssays* **25**: 1147–9.
34. **Barreiro LB**, **Tailleux L**, **Pai AA**, **Gicquel B**, et al. 2012. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *Proc Natl Acad Sci USA* **109**: 1204–9.
35. **Welbourn S**, **Miyagi E**, **White TE**, **Diaz-Griffero F**, et al. 2012. Identification and characterization of naturally occurring splice variants of SAMHD1. *Retrovirology* **9**: 86.
36. **Pauls E**, **Jimenez E**, **Ruiz A**, **Permanyer M**, et al. 2013. Restriction of HIV-1 replication in primary macrophages by IL-12 and IL-18 through the upregulation of SAMHD1. *J Immunol* **190**: 4736–41.
37. **Hollenbaugh JA**, **Gee P**, **Baker J**, **Daly MB**, et al. 2013. Host factor SAMHD1 restricts DNA viruses in non-dividing myeloid cells. *PLoS Pathog* **9**: e1003481.
38. **Rehwinkel J**, **Maelfait J**, **Bridgeman A**, **Rigby R**, et al. 2013. SAMHD1-dependent retroviral control and escape in mice. *EMBO J* **32**: 2454–62.
39. **Zhang R**, **Bloch N**, **Nguyen LA**, **Kim B**, et al. 2014. SAMHD1 restricts HIV-1 replication and regulates interferon production in mouse myeloid cells. *PLoS One* **9**: e89558.
40. **Hamilton CE**, **Papavasiliou FN**, **Rosenberg BR.** 2010. Diverse functions for DNA and RNA editing in the immune system. *RNA Biol* **7**: 220–8.
41. **Roberts SA**, **Lawrence MS**, **Klimczak LJ**, **Grimm SA**, et al. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* **45**: 970–6.
42. **Rosenberg BR**, **Hamilton CE**, **Mwangi MM**, **Dewell S**, et al. 2011. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nat Struct Mol Biol* **18**: 230–6.
43. **Athanasiadis A**, **Rich A**, **Maas S.** 2004. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* **2**: 2144–58.
44. **Kim DDY**, **Kim TTY**, **Walsh T**, **Kobayashi Y**, et al. 2004. Widespread RNA editing of embedded Alu elements in the human transcriptome. *Genome Res* **14**: 1719–25.
45. **Melcher T**, **Maas S**, **Herb A**, **Sprengel R**, et al. 1996. A mammalian RNA editing enzyme. *Nature* **379**: 460–4.
46. **Valente L**, **Nishikura K.** 2005. ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation. *Prog Nucleic Acid Res Mol Biol* **79**: 299–338.
47. **Fritz EL**, **Rosenberg BR**, **Lay K**, **Mihailovic A**, et al. 2013. A comprehensive analysis of the effects of the deaminase AID on the transcriptome and methylome of activated B cells. *Nat Immunol* **14**: 749–55.
48. **Doria M**, **Neri F**, **Gallo A**, **Farace MG**, et al. 2009. Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic Acids Res* **37**: 5848–58.
49. **Ho YC**, **Shan L**, **Hosmane NN**, **Wang J**, et al. 2013. Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**: 540–51.
50. **Shan L**, **Siliciano RF.** 2013. From reactivation of latent HIV-1 to elimination of the latent reservoir: The presence of multiple barriers to viral eradication. *BioEssays* **35**: 544–52.
51. **Sharkey M**, **Babic DZ**, **Greenough T**, **Gulick R**, et al. 2011. Episomal viral cDNAs identify a reservoir that fuels viral rebound after treatment interruption and that contributes to treatment failure. *PLoS Pathog* **7**: e1001303.
52. **Kim S**, **Lee S**, **Shin J**, **Kim Y**, et al. 2011. Human cytomegalovirus microRNA miR-US4-1 inhibits CD8(+) T cell responses by targeting the aminopeptidase ERAP1. *Nat Immunol* **12**: 984–91.
53. **Yao Y**, **Nair V.** 2014. Role of virus-encoded microRNAs in avian viral diseases. *Viruses* **6**: 1379–94.
54. **Huang WT**, **Lin CW.** 2014. EBV-encoded miR-BART20-5p and miR-BART8 inhibit the IFN-gamma-STAT1 pathway associated with disease progression in nasal NK-cell lymphoma. *Am J Pathol* **184**: 1185–97.
55. **Wong SK**, **Sato S**, **Lazinski DW.** 2001. Substrate recognition by ADAR1 and ADAR2. *RNA* **7**: 846–58.
56. **Tian N**, **Yang Y**, **Sachsenmaier N**, **Muggenhumer D**, et al. 2011. A structural determinant required for RNA editing. *Nucleic Acids Res* **39**: 5669–81.
57. **Jacobs MM**, **Fogg RL**, **Emeson RB**, **Stanwood GD.** 2009. ADAR1 and ADAR2 expression and editing activity during forebrain development. *Dev Neurosci* **31**: 223–37.
58. **Wahlstedt H**, **Daniel C**, **Enstero M**, **Ohman M.** 2009. Large-scale mRNA sequencing

determines global regulation of RNA editing during brain development. *Genome Res* **19**: 978–86.

59. **Blanc V**, **Xie Y**, **Luo JY**, **Kennedy S**, et al**.** 2012. Intestine-specific expression of Apobec-1 rescues apolipoprotein B RNA editing and alters chylomicron production in Apobec1(−/−) mice. *J Lipid Res* **53**: 2643–55.
60. **Hashimoto Y**, **Matsuoka M.** 2014. A mutation protective against Alzheimer's disease renders amyloid beta precursor protein incapable of mediating neurotoxicity. *J Neurochem* **130**: 291–300.
61. **Tu ZD**, **Keller MP**, **Zhang CS**, **Rabaglia ME**, et al**.** 2012. Integrative analysis of a cross-loci regulation network identifies App as a gene regulating insulin secretion from pancreatic islets. *PLoS Genet* **8**: e1003107.
62. **Yin H**, **Xue W**, **Chen S**, **Bogorad RL**, et al**.** 2014. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* **32**: 551–3.
63. **Sander JD**, **Joung JK.** 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**: 347–55.
64. **Hassan MA**, **Melo MB**, **Haas B**, **Jensen KD**, et al**.** 2012. De novo reconstruction of the *Toxoplasma gondii* transcriptome improves on the current genome annotation and reveals alternatively spliced transcripts and putative long non-coding RNAs. *BMC Genomics* **13**: 696.
65. **Robertson G**, **Schein J**, **Chiu R**, **Corbett R**, et al**.** 2010. De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**: 909–12.
66. **Bernstein BE**, **Meissner A**, **Lander ES.** 2007. The mammalian epigenome. *Cell* **128**: 669–81.
67. **Cortijo S**, **Wardenaar R**, **Colome-Tatche M**, **Gilly A**, et al**.** 2014. Mapping the epigenetic basis of complex traits. *Science* **343**: 1145–8.
68. **Malireddy S**, **Kotha SR**, **Secor JD**, **Gurney TO**, et al**.** 2012. Phytochemical antioxidants modulate mammalian cellular epigenome: implications in health and disease. *Antioxid Redox Signal* **17**: 327–39.
69. **Vollmers C**, **Schmitz RJ**, **Nathanson J**, **Yeo G**, et al**.** 2012. Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metab* **16**: 833–45.
70. **Whitehead J**, **Pandey GK**, **Kanduri C.** 2009. Regulation of the mammalian epigenome by long noncoding RNAs. *Biochim Biophys Acta* **1790**: 936–47.
71. **Bird A.** 2002. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**: 6–21.
72. **Gelfman S**, **Cohen N**, **Yearim A**, **Ast G.** 2013. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res* **23**: 789–99.
73. **Lister R**, **Pelizzola M**, **Dowen RH**, **Hawkins RD**, et al**.** 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315–22.
74. **Laurent L**, **Wong E**, **Li G**, **Huynh T**, et al**.** 2010. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**: 320–31.
75. **Gelfman S**, **Ast G.** 2013. When epigenetics meets alternative splicing: the roles of DNA methylation and GC architecture. *Epigenomics* **5**: 351–3.
76. **Barbosa-Morais NL**, **Irimia M**, **Pan Q**, **Xiong HY**, et al**.** 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**: 1587–93.
77. **Anastasiadou C**, **Malousi A**, **Maglaveras N**, **Kouidou S.** 2011. Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers. *DNA Cell Biol* **30**: 267–75.
78. **Tao Y**, **Xi S**, **Briones V**, **Muegge K.** 2010. Lsh mediated RNA polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation. *PLoS One* **5**: e9163.
79. **Vignali M**, **Hassan AH**, **Neely KE**, **Workman JL.** 2000. ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol* **20**: 1899–910.
80. **Chen W**, **Luo L**, **Zhang L.** 2010. The organization of nucleosomes around splice sites. *Nucleic Acids Res* **38**: 2788–98.
81. **McVicker G**, **van de Geijn B**, **Degner JF**, **Cain CE**, et al**.** 2013. Identification of genetic variants that affect histone modifications in human cells. *Science* **342**: 747–9.
82. **Gaffney DJ**, **McVicker G**, **Pai AA**, **Fondufe-Mittendorf YN**, et al**.** 2012. Controls of nucleosome positioning in the human genome. *PLoS Genet* **8**: e1003036.
83. **Kumar SV**, **Wigge PA.** 2010. H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. *Cell* **140**: 136–47.
84. **Wolff GL**, **Kodell RL**, **Moore SR**, **Cooney CA.** 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J* **12**: 949–57.
85. **Waterland RA**, **Jirtle RL.** 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* **23**: 5293–300.
86. **Foll M**, **Gaggiotti O.** 2006. Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**: 875–91.
87. **Rosenberg NA.** 2006. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet* **70**: 841–7.
88. **Cann HM**, **de Toma C**, **Cazes L**, **Legrand MF**, et al**.** 2002. A human genome diversity cell line panel. *Science* **296**: 261–2.