

Review

Cloudy with a Chance of Insights: Context Dependent Gene Regulation and Implications for Evolutionary Studies

Elisa Buchberger ^{1,†}, Micael Reis ^{1,†}, Ting-Hsuan Lu ^{1,2} and Nico Posnien ^{1,*} 

¹ Dpt. of Developmental Biology, Göttingen Center for Molecular Biosciences (GZMB), University Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany

² International Max Planck Research School for Genome Science, Am Fassberg 11, 37077 Göttingen, Germany

* Correspondence: nposnie@gwdg.de

† Equally contributed.

Received: 24 May 2019; Accepted: 26 June 2019; Published: 28 June 2019



Abstract: Research in various fields of evolutionary biology has shown that divergence in gene expression is a key driver for phenotypic evolution. An exceptional contribution of *cis*-regulatory divergence has been found to contribute to morphological diversification. In the light of these findings, the analysis of genome-wide expression data has become one of the central tools to link genotype and phenotype information on a more mechanistic level. However, in many studies, especially if general conclusions are drawn from such data, a key feature of gene regulation is often neglected. With our article, we want to raise awareness that gene regulation and thus gene expression is highly context dependent. Genes show tissue- and stage-specific expression. We argue that the regulatory context must be considered in comparative expression studies.

Keywords: gene expression; gene regulation; evolution; allele specific expression; expression quantitative trait loci (eQTL); RNAseq; ChIPseq; chromatin; Assay for Transposase-Accessible Chromatin using sequencing (ATACseq); genotype–phenotype map

1. Introduction

Living organisms are uniquely characterized by their appearance, their function, and their interaction with the environment. The information about these features is provided in the genome which is packed into the nucleus of each cell (see Figure 1A). Various disciplines of biological and medical research aim at understanding how the genomic information is transformed into organismic functionality. Proteins and peptides are the molecules that accomplish manifold tasks in an organism, such as orchestrating its development [1], providing energy through metabolism [2,3], protection via immune responses [4,5], and processing environmental information in the nervous system [6,7]. Protein and peptide sequences are encoded in gene regions of the genome. Genes are transcribed into ribonucleic acid (RNA) molecules that serve as templates for the translation machinery that eventually synthesizes functional proteins. This process, called gene expression, is thus crucial for every living organism.

Since the identification of deoxyribonucleic acid (DNA) as genetic material in 1944 [8] a major focus in Evolutionary Biology and Quantitative Genetics has been to reveal the connections between variation in DNA sequences and phenotypic differences observed among organisms (i.e., the genotype–phenotype map) [9–12]. If causative genetic variation is identified in protein coding sequences it is straightforward to directly link these differences to changes in protein function [13–16]. However, if causative genetic variation is present in intergenic or intronic (i.e., non-coding) sequences it is less intuitive to infer direct

links between the observed variation and phenotypic differences. Since these non-coding regions may be important regulatory sequences it is conceivable to connect genetic variation in such regions with differential gene expression. With the advent of efficient and affordable sequencing technologies (next generation sequencing, NGS) it became feasible to study gene expression on a genome wide scale [17]. Since these technologies also provide the opportunity to obtain such data in plant and animal systems beyond well-established genetic models, gene expression has extensively been used as proxy for genetic variation to gain insights into phenotypic evolution [18,19].

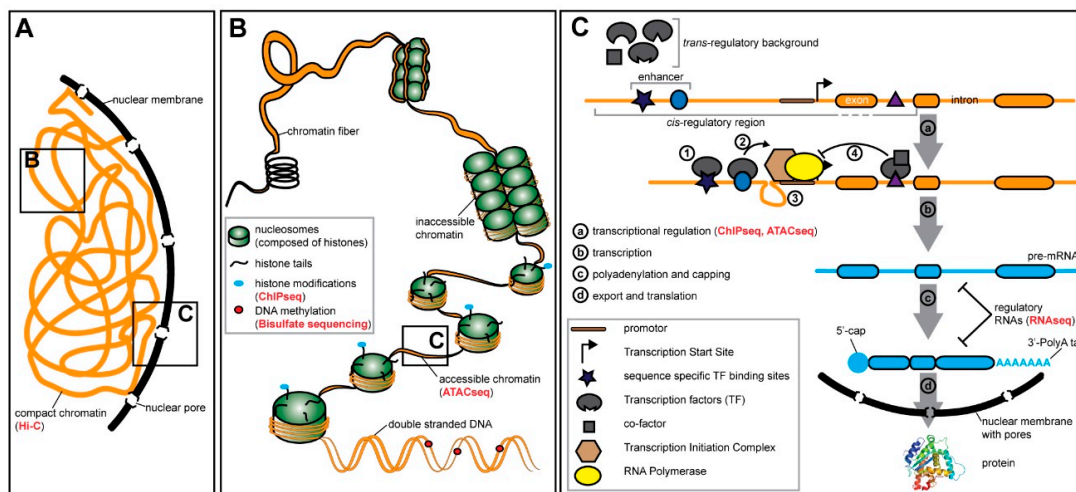


Figure 1. Gene expression is regulated on various levels. (A) The DNA is compressed in the nucleus of the cell. (B) The DNA in the nucleus is compressed by binding of histone proteins. The chromatin contains easily accessible euchromatin regions and highly compact and inaccessible heterochromatin regions. The status of the chromatin is influenced by post-translational histone modifications. Gene expression is modulated by the chromatin state and DNA modifications, such as methylations. (C) Key steps of gene expression (a–d). Transcription factors (TFs) bind to the DNA at specific sequences (1). TF binding activates the transcription initiation complex (2) through conformation changes (looping) of the DNA (3). TFs can also repress transcription, for instance by binding of a co-factor (4). Next generation sequencing (NGS)-based methods that can be applied to study certain aspects of gene regulation are mentioned in red in brackets. See Table 1 for an overview of the methods mentioned here.

In this review we will first summarize findings illustrating the importance of gene expression divergence in phenotypic evolution for various morphological, behavioral, physiological, and life-history traits. Next, we will present current approaches aiming at understanding genome-wide patterns of gene expression divergence as well as the underlying molecular mechanisms. We will review various mechanisms underlying gene regulation and we will highlight how they facilitate context dependent gene expression. We argue that gene expression and gene regulation evolve in a highly context dependent manner and we will suggest to consider that knowledge to improve the efficiency of comparative gene expression studies.

2. Gene Expression Divergence Affects Phenotypic Evolution

Changes in gene expression have been linked to variation in many phenotypes. In the last years, there has been an increase in the number of ecological and evolutionary studies using transcriptomics to understand how environment and different life strategies affect gene expression [20,21]. Most examples found in the literature connecting genetic variation affecting gene expression with phenotypes are based on studying simple morphological traits, such as the evolution of trichome patterns in *Drosophila* or differences in body coloration. For instance, a clear link between changes in the regulatory region of the *shavenbaby* gene and the evolution of trichome patterns across *Drosophila* species has been

established [22,23]. Similarly, individual nucleotide polymorphisms in the *ebony* [24] and *yellow* genes [25] underlie natural differences in body and wing pigmentation, respectively, in *Drosophila*. Divergence in fur coloration in mice has been shown to be regulated by differences in developmental expression of the gene *agouti* [26,27]. Moreover, the stripe pattern of cichlid fishes is associated with differential expression of the gene *agouti-related peptide 2 (agrp2)* [28].

Besides these classical traits, also more complex traits are being studied. In *Drosophila*, the shape of male genitalia evolves rapidly, contributing to speciation processes. Divergence in the expression of the *tartan* gene has recently been shown to contribute to interspecific differences between *D. mauritiana* and *D. simulans* [29]. Another study has shown that a single nucleotide change in the *cis*-regulatory region of *scute* has pleiotropic effects by affecting genitalia bristle and sex comb sensory teeth number simultaneously [30]. Hence, gene expression divergence is a major driver of the evolution of morphological traits.

Recently it has been argued that the molecular architecture of differences in behavioral traits may be simpler than previously anticipated. For instance, a complex behavior such as sociality in bees has been shown to be clearly associated with differential expression of the gene *syntaxin1a*, since higher expression of this gene is directly correlated with a social life style [31]. Similarly, differences in parental care between the promiscuous deer mouse (*Peromyscus maniculatus bairdii*) and its sister species, the monogamous old-field mouse (*P. polionotus subgriseus*) is influenced by differential expression of the gene *vasopressin* [32]. These examples impressively demonstrate that the evolution of behavioral traits is associated with divergence in gene expression.

Many studies exploring the molecular basis of the evolution of physiological and life-history traits followed by functional validation have confirmed an underlying polygenic architecture [12,33,34]. Nevertheless, few studies reached the resolution to narrow down genetic variation to the level of individual loci. A recent study in European aspen (*Populus tremula*) has shown that expression divergence of a single gene (*PtFT2*) is responsible for 65% of the differences in timing of bud set [35]. Other studies similarly identified mutations in *cis*-regulatory regions causing gene expression divergence which ultimately affects an organism's physiological response to the environment. For example, a 2 bp deletion in the promoter region of the gene *ERG28* in *Saccharomyces cerevisiae* results in reduced expression associated with resistance to an antifungal drug [36]. Similarly, an indel in the 3'UTR of *MtnA* that shows signatures of selection, causes a 4-fold difference in gene expression and confers resistance to oxidative stress in natural populations of *D. melanogaster* [37].

In summary, genetic variation associated with the evolution of phenotypic traits such as morphology, behaviour, life history, and physiology often affect gene expression. Therefore, gene expression divergence is a major driver for phenotypic evolution.

3. Gene Expression Divergence Reflects Divergent Gene Regulatory Mechanisms

Since gene expression divergence is often linked to phenotypic evolution, many comparative studies employ gene expression as intermediate phenotype to link genetic variation to trait divergence. However, besides using gene expression to establish genotype–phenotype maps, the availability of high throughput methods to survey genome wide expression levels also allows to study global patterns of gene expression divergence. It has for instance been shown that divergence in gene expression is pervasive among populations in *Drosophila* [38], yeast [39], or in fish [40]. It is therefore likely that genetic variants affecting gene expression segregate in natural populations and can be selected for [41]. Also, interspecific gene expression data across seven *Drosophila* species has been used in modelling approaches integrating fitness estimates and a phylogenetic framework to reveal that expression divergence shows signatures of directional selection [42]. Additionally, interspecific gene expression comparisons contributed to a better understanding of biological phenomena such as sex-biased gene expression [43] or expression variation of duplicated genes [44]. Therefore, comparative gene expression studies revealed a high level of variation in gene expression within and among species.

The accumulation of comparative genome wide expression data triggered a strong interest in unravelling the molecular and evolutionary mechanisms underlying divergence in gene expression itself. Most of our current mechanistic understanding of gene expression divergence is based on work in genetic model systems that are tractable for genetic crosses. Two main methods have been employed extensively in recent years, i.e., expression quantitative trait loci (eQTL) mapping and allele-specific expression studies (ASE). eQTL studies are basically QTL or genome wide association (GWAS) studies aiming at identifying causative loci responsible for gene expression variation. Conceptionally, this method assumes that the level of gene expression can be treated as a quantitative trait [45–47]. Therefore, normal QTL or association mapping methodology can be applied to reveal genomic variants associated with expression divergence. eQTL studies supported the fundamental observation that gene expression is indeed highly variable across individuals and heritability estimates support the contribution of a genetic component [48].

While eQTL studies reveal genomic loci or individual single nucleotide polymorphisms (SNPs) associated with expression difference, ASE studies in F1 hybrids represent a powerful approach to gain mechanistic insights into differential gene expression [49]. The analysis of gene expression between homozygous parents (closely related species or populations of the same species) and the allele specific expression in their heterozygous F1 offspring allows distinguishing whether a gene is differentially expressed due to changes in its own regulatory region (*cis*-regulatory divergence) or due to changes somewhere else in the genome (*trans*-regulatory divergence) [50,51]. *cis*-regulatory divergence is inferred if two different alleles of a given gene have a major impact on its allele-specific expression in the homogenous *trans*-regulatory background of the F1 hybrid. *trans*-regulatory divergence is inferred if a gene is differentially expressed between two parental individuals, but the contribution of the two alleles in the hybrid background is the same. The most consistent observation in ASE studies is that *cis*-regulatory divergence seems to be prevalent in intra- as well as interspecific comparisons [52–55]. Exceptions have been observed for instance for comparisons between the cosmopolitan fly species *D. melanogaster* and the closely related specialist species *D. sechellia* [56]. In all mentioned ASE studies, a major impact of a combination of *cis*- and *trans*-divergence has been observed, strongly supporting the notion that gene regulation is complex and thus can evolve in complex patterns.

ASE studies alone do not allow revealing genetic variants associated with gene expression divergence. However, one of the most likely explanation for *cis*-divergence effects is sequence variation in the regulatory region (i.e., promoters or enhancers) of the differentially expressed gene. Indeed, in putative regulatory regions of genes showing *cis*-regulatory divergence increased levels of sequence divergence have been found in yeast [57], *Arabidopsis thaliana* [58], maize [59], and *Drosophila* [56,60]. A combination of ASE and SNP data obtained from lymphoblastoid cell lines from the 1000 Genomes Project further strongly suggests that genetic variation is a common explanation for allele-specific gene expression [61]. Hence, the combination of ASE and eQTL studies provide exceptional insights into gene expression divergence.

In summary, comparative gene expression studies are extensively used to establish genotype-phenotype relationships and to reveal global patterns of expression variation. eQTL and ASE furthermore represent excellent approaches to gain mechanistic insights into gene expression divergence.

4. Gene Expression and Gene Regulation are Highly Context Dependent

Many of the above-mentioned exciting fundamental insights into gene expression divergence are based on studies in entire organisms and adult stages. However, it is broadly accepted that gene expression is strongly context dependent with a major impact for instance of the developmental stage and the tissue [62]. Focusing comparative gene expression studies on a few stages of an organism's life history and a combined view of usually complex tissue compositions underestimates this important aspect of gene expression and gene regulation. Since the molecular mechanisms underlying context dependent gene regulation are being revealed these days, we will summarize key features of the gene regulation machinery and highlight how they facilitate context dependent gene expression.

4.1. Pre-transcriptional Regulation—Chromatin States and Methylation

The first regulatory mechanisms are at play on the level of genome organisation. Compressed DNA in the nucleus forms a tertiary structure (Figure 1A) that can be studied in detail by an NGS based chromosome conformation capture method called Hi-C [63] (Table 1). Hi-C applied in various bilaterians revealed one fundamental characteristic of the genome: Some regions of the genome interact consistently more often than other regions [64–66]. These topologically associating domains (TADs) have been shown to influence gene expression. For instance, the famous temporal and spatial collinearity of *Hox* gene expression in the developing vertebrate limb has been associated with the location of the *HoxD* cluster in a gene desert that lies between two adjacent TADs [67]. The application of Hi-C in different human primary blood cell types showed that these interactions are highly cell-type specific [68]. How the three-dimensional organization of the genome exactly influences gene regulation has just started to be revealed and represents an active and exciting field of research.

Highly transcribed regions of the genome (i.e., euchromatin) are—in contrast to the condensed heterochromatin—usually depleted of nucleosomes (Figure 1B). These nucleosome free regions can be detected on a genome wide scale using NGS based methods such as ATACseq (Assay for Transposase-Accessible Chromatin using sequencing) [69] (Table 1). Recent application of ATACseq on single cells originating from 13 different mouse tissues [70] and from three stages of *Drosophila* embryonic development [71] revealed clear signatures of cell type and stage specific chromatin accessibility states. The chromatin state and thus DNA accessibility is influenced by cell type specific modifications of histone proteins [72,73], the subunits of nucleosomes, and is clearly linked to gene regulation [74,75]. Even if DNA is accessible, the transcription of genes can be modulated by DNA methylation, i.e., the addition of a methyl group to cytosines. DNA methylation has been associated with gene repression [76,77] and recent data has shown that transcription factors can integrate methylation patterns to refine gene regulation [78]. Since the methylation is highly dynamic, for instance throughout cellular differentiation [79], it facilitates context dependent gene regulation.

Chromatin accessibility and genome architecture is also regulated by a variety of non-coding RNA molecules, which are transcribed, but not translated into proteins. Long non-coding RNAs (lncRNAs) localized in the nucleus can directly affect chromatin architecture [80]. Intriguingly, lncRNA–protein interactions are tissue specific since they were observed in mouse placenta cells, but not in liver tissue [81]. lncRNAs have also been implicated in transcriptional activation since they mediate active histone marks [82] and stabilize enhancer-promotor interactions [83]. Micro RNAs (miRNAs), another group of non-coding RNAs, have been shown to directly modulate histone modifications and thus the chromatin accessibility to allow transcription of target genes [84]. Since miRNAs are highly tissue specifically expressed [85–87], these molecules provide an excellent mechanism to facilitate tissue specific chromatin accessibility.

In summary, the extensive diversity of epigenetic modifications, which are further modulated by non-coding RNAs regulate differential DNA accessibility and thus provide a rich cellular repertoire to control gene expression pre-transcriptionally.

4.2. Transcriptional Regulation—Transcription Factors and Cis-Regulatory Elements

Once the chromatin is accessible for proteins, gene expression is directly regulated by protein–DNA interactions [88,89] (Figure 1C). Transcription factors are proteins with dedicated DNA-binding domains and their sequence specific binding fosters or represses gene expression. A classic example for context-specific gene regulation via transcription factors is the development of different neuronal subtypes in the *Drosophila* central nervous system. Initially, all neuronal precursor cells, the neuroblasts, contain generic neuronal transcription factors [90]. The unique identity of each neuroblast is further specified by spatial and temporal cues. Different neuronal subtypes are defined by the expression of temporally restricted transcription factors [91] and the regional identity of neuroblasts is regulated by the expression of spatially restricted transcription factors [92]. Therefore, context dependent gene

regulation can be achieved on the level of the presence of transcription factors which are expressed cell and time specifically.

Besides the transcription factors itself, the nature of the DNA sequences they bind to plays a major role in gene regulation. These *cis*-regulatory regions can be subdivided based on their location relative to the respective gene locus (Figure 1C). Promoters lie directly upstream of the transcription start site and general transcription factors bind there as part of the transcription initiation complex [93,94]. Enhancers are *cis*-regulatory sequences that are located further away up- and downstream of the transcription start site. They are composed of distinct sequence motifs that are specifically recognized by certain transcription factors. Transcription factors bound to enhancers facilitate the assembly and activation of the transcription initiation complex at the promoter [95,96]. Although we focus here on enhancers, many of the discussed aspects apply to other elements, such as silencers and insulators as well. Enhancers are highly modular [97,98], as exemplified by the regulation of the pair-rule gene *even skipped* (*eve*) during segmentation in the *Drosophila* embryo. The seven stripes of *eve* expression are spatially defined by five enhancers with each of them being responsible for an individual stripe or a pair of stripes [99,100]. Therefore, the modular nature of enhancers provides a source for context-dependent activation (and repression) of genes.

The interaction of transcription factors and *cis*-regulatory elements can be further diversified by the interaction of transcription factors with co-factors that are expressed in a temporally and spatially defined manner to modulate for instance their capacity to bind to regulatory regions (Figure 1C). One excellent example for the context dependence of gene regulation achieved via the spatial availability of co-factors has been shown in the developing wing disc of *Drosophila*. During wing development, the transcription factor Pannier (Pnr) can act as an activator in some regions, while the presence and binding of its co-factor U-shaped (Ush) transforms it into a transcriptional repressor in adjacent regions [101–103]. The importance of transcriptional co-factors has also been shown on a genome wide scale. For instance, the two transcription factors CLOCK (CLK) and CYCLE (CYC), which are core components of the circadian clock in flies, are broadly expressed. However, the tissue specific response to the circadian clock is defined by the action of co-factors, which modulate the DNA binding capacities of these two transcription factors in a tissue specific manner [104]. The modulation of protein-DNA interactions by co-factors bound to transcription factors thus provides an additional mechanism to ascertain context-dependent gene expression.

In summary, the interaction of spatially and temporarily expressed transcription factors, with modular regulatory DNA sequences specifies the unique transcriptional landscape of a developing cell or cell groups.

4.3. Post-transcriptional Regulation—RNA Modifications and Regulatory RNA Molecules

Apart from the regulation of transcription itself, the transcriptional outcome can be fine-tuned on the level of the messenger RNA. For instance, post-transcriptional modifications, such as polyadenylation and capping influence mRNA export, stability and translation efficiency [105–107]. Differential splicing of primary transcripts allows enlarging the repertoire of proteins to be translated from a limited number of primary RNAs. Splicing is mediated by a specific protein-complex [108] and it has been shown that tissue and cell type specific patterns of splicing factor expression recapitulate the extent of alternative spliced transcripts present in the respective tissue [109].

Post-transcriptional gene regulation is also mediated by regulatory RNA molecules, which can be involved in negative gene regulation via the RNA interference (RNAi) pathway (e.g., miRNA) [110] or they are part of RNA-protein complexes (e.g., lncRNA) where they influence gene regulation on various levels [111]. lncRNAs present in the cytoplasm also influence mRNA stability [112,113] and they can protect mRNA against targeted degradation by trapping miRNAs in a sponge-like mechanisms [114]. Regulatory RNAs play a major role during development [111,115] and their expression has been shown to be cell type specific [116,117]. Context dependent gene regulation can thus be mediated

post-transcriptionally by differences in generic RNA modification programs (e.g., splicing) or by the action of regulatory RNA molecules (miRNA, lncRNA).

Table 1. Next generation sequencing techniques used for studying gene expression and gene regulation in evolutionary studies. Methods labelled with * require a reference genome.

Method	Key information
RNAseq	<p>Summary: RNA is isolated and reverse transcribed into cDNA for library preparation and sequencing.</p> <p>Practical considerations: The most common protocol uses oligo-dT primers to enrich for polyadenylated RNAs for reverse transcription of processed mRNA [17] and the majority of lncRNAs [118]. Alternative protocols use total RNA and ribosome depletion prior to reverse transcription with random oligos to obtain other RNA molecules (e.g., immature mRNA, miRNA, and siRNA) [119]. For small RNA enrichment several commercial kits are available to select for molecule sizes less than 30 nucleotides [120].</p> <p>Applications: Transcriptome generation for gene annotation including alternative isoforms (paired-end sequencing) and differential gene expression analysis between different samples (e.g., tissues, experimental conditions, populations of the same species or even species showing different phenotypes) [18–21,121]. RNAseq is also a useful tool for miRNA profiling and annotation [122] as well as differential expression of lncRNAs [123].</p> <p>Single cell application: [124–126]</p>
ATACseq*	<p>Summary: Accessible chromatin regions which are not condensed by histones, are digested with a genetically modified transposase (Tn5). Nucleotide overhangs (tagmentation) are utilized for specific adapter ligation during the library preparation and sequencing [69,127]. This method substituted previous ones such as DNase-seq and FAIREseq, due to its simplicity and effectiveness.</p> <p>Practical considerations: Usually the protocol should be done with fresh tissue and a defined number of nuclei/cells (e.g 500–50,000 for mammalian tissues [127]) that have to be estimated prior to tagmentation. These technical aspects limit the number of samples that can be processed simultaneously. However, protocols were successfully applied to frozen tissue [128].</p> <p>Applications: ATACseq is commonly used to complement RNAseq data to identify potential regulatory regions (enhancers) [129]. ATACseq can also be used to evaluate chromatin structure dynamics and epigenetic changes by providing information about histone position as well as a complementary approach to ChIPseq to characterize transcription factor and repressor (e.g., CTCF) occupancies [69].</p> <p>Single cell application: [130,131]</p>
ChIPseq*	<p>Summary: DNA bound proteins (e.g., transcription factors, histones) are crosslinked and the chromatin is digested with restriction enzymes. Antibodies specific for the DNA-binding protein are used to isolate Protein-DNA fragments. After reversal of the crosslink and dissociation of the DNA short read sequencing libraries are prepared [132,133].</p> <p>Practical considerations: This technique relies on previous knowledge about the DNA-binding proteins and available antibodies.</p> <p>Applications: ChIPseq is commonly used to generate genome wide data on protein-DNA interactions, mainly to determine transcription factor binding sites and their binding dynamics [134]. It has been used also to estimate histone modifications and nucleosome position between different species [72].</p> <p>Single cell application: [135]</p>
Hi-C*	<p>Summary: DNA-binding proteins and chromatin are covalently crosslinked with formaldehyde and digested with a restriction enzyme. The resulting fragments are ligated to create chimeric molecules of DNA which are further isolated for library preparation and sequencing [136].</p> <p>Practical considerations: Hi-C relies on restriction enzyme recognition sites which can create bias due to their heterogeneous distribution in the genome [137]. Alternative methods used DNase I [138] or microcococcus digestion [139] to overcome that issue.</p> <p>Applications: Hi-C is commonly used to identify global patterns of 3D genome conformation. Additionally, this method allows exploring how interactions between different chromosomal regions can affect gene regulation. The impact of chromatin topology on gene expression between species has been studied [64,66,140].</p> <p>Single cell application: [141]</p>
BSseq	<p>Summary: DNA is treated with sodium bisulfite to deaminate cytosine bases into uracil (thymine after PCR) while methyl-cytosine bases are not affected [142]. The treated DNA is then digested for library preparation and sequencing [143].</p> <p>Practical considerations: The deamination reaction usually has high yield, but small variations can create significant bias in the estimation of global methylation patterns [144]. Since cytosine is converted into thymine, the sequence complexity is reduced, and the strands are no longer complementary causing potential problems with the alignments. However, dedicated software has been developed to deal with the challenging BSseq data analysis (reviewed in [145]).</p> <p>Applications: This method is used to obtain genome wide patterns of DNA methylation which is an important epigenetic modification typically associated with gene expression repression [143]. In recent years, this method has been extensively applied to ecological and evolutionary studies [144,146].</p> <p>Single cell application: [147,148]</p>

5. The Evolution of Gene Expression and Gene Regulation is Context Dependent

So far, we established that variation in gene expression is pervasive within and across species and that it is a major driver of phenotypic divergence. Furthermore, we showed that the different levels of the gene regulation machinery facilitate context dependent gene expression. However, few comparative expression studies have specifically tested for the context dependency of gene expression variation across species. A study of six homologous organs in nine mammals and one bird, for instance, showed that gene expression evolves at different speeds in different tissues as well as in different lineages. While gene expression was stable in the nervous system, it evolved more rapidly in testes. Similarly, gene expression variation was less pronounced in rodents compared to apes [149]. Comparative studies have also been employed to assess the impact of developmental stages on the evolution of gene expression. The analysis of expression data from various developmental stages in different vertebrates revealed the pharyngula stage to be most constraint (i.e., most similar) [150,151]. Intriguingly, a similar analysis restricted to the developing brain, instead of entire embryos, identified a stage of high conservation of gene expression much later just before birth [152]. These examples clearly demonstrate that the gene expression context, such as the type of tissue or the developmental stage, poses constraints on the overall evolvability of gene expression. Thus, the context in which gene expression variation is studied will dramatically affect the results.

A comprehensive understanding of the regulatory mechanisms underlying the context dependent evolution of gene expression is still missing to date. Indeed, few eQTL and ASE studies specifically compared findings across different tissues or stages to reveal context dependent regulatory mechanisms. A comparison between mouse embryonic and adult tissue has shown that many more distal (“*trans*”) eQTLs were found in adults compared to the investigated embryonic stage [153]. Similarly, the analysis of sexually dimorphic gene expression in different organs in intercrosses of two inbred mouse strains revealed tissue specific eQTL regions, suggesting that expression differences between sexes are regulated by tissue specific regulatory elements [154]. ASE studies have revealed different contributions of *cis*- and *trans*-divergence (or combinations thereof) by comparing differently aged flies [155] and when data from entire fly bodies was compared to heads only [54,56]. Moreover, a recent ASE study using tissue specific data for Malpighian tubules of different *D. melanogaster* populations further supports the need for more defined analyses [156]. In the light of context dependent gene regulation these first results call for an integration of stage or tissue specific aspects of gene expression in eQTL and ASE studies in order to reveal whether patterns observed so far will hold true across highly variable regulatory environments.

Indeed, the combination of various genome wide datasets for a highly context specific cellular system has already contributed to exciting insights into the impact of natural genetic variation on the different levels of the gene regulation machinery (see Figure 1). As part of the HapMap2 [157] and 1000 Genomes Project [61] lymphoblastoid cell lines were established from hundreds of individuals and subjected to genome sequencing, providing a solid basis for association studies for various regulatory traits in combination with expression variation. These studies revealed SNPs affecting all levels of gene regulation including genome organization [158], chromatin accessibility [159], histone modifications, RNA-Polymerase II occupancy, and eventually gene expression [160,161]. About 65% of the eQTLs (i.e., variation in gene expression) are associated with histone modifications and chromatin accessibility [162], suggesting that additional mechanisms must contribute to gene expression divergence. Post-transcriptional processes such as mRNA splicing are excellent candidates since individual SNPs have been associated with differences in splicing [162,163]. Since the spliceosome is already assembled during ongoing transcription, the chromatin state and the transcription rate can influence splicing events [164]. Interestingly, many identified SNPs affect different regulatory mechanisms simultaneously. For instance, genetic variants that confer higher transcription factor binding affinity are also associated with an increase in active histone marks [160], suggesting a tight causal link between transcription factor binding and histone modifications. Another link has been established between natural variation in epigenetic methylation patterns and gene expression,

by showing that the same genetic variant is associated with variation in gene expression and the methylation of a CpG island close to the respective gene locus [165]. Variation in gene expression, gene regulation and methylation are therefore tightly linked.

In the light of recent findings that the rate of gain and loss of active enhancer elements in five closely related *Drosophila* species is relatively high [166], it is conceivable that natural genetic variation very quickly affects gene expression on various levels ranging from transcription factor binding to histone modification and chromatin accessibility. Additionally, a comparative study of methylation in promotor regions of primates has shown that methylated CpG islands are characterized by a higher mutation rate and that the loss of CpG islands in humans is most likely driven by methylation in sperm [167]. The observation that many regulatory traits are functionally linked and thus similarly affected by the same SNPs may explain why natural variation in gene expression is pervasive and can drive phenotypic diversification. Since many of the genome wide sequencing methods (Table 1) are readily applicable in various systems, more studies in the future will allow establishing clear links between natural variation on various levels of gene regulation and tissue-, cell-, and stage-specific evolution of gene expression.

Since the complexity of regulatory interactions is highly context dependent, tissue-, stage-, or species-specific constraints may also be imposed by the gene regulatory networks in which certain gene products act. In the yeast *Saccharomyces cerevisiae* for instance, endogenous processes like cell cycle progression are regulated by highly complex networks, while simpler networks regulate processes that result from external stimuli (e.g., stress response) [168,169]. Highly connected genes in more complex gene regulatory networks tend to be more conserved within and across *Drosophila* species [170]. Similarly, gene regulatory networks are extensively rewired throughout development. For instance, the gene regulatory network underlying *Drosophila* trichome formation at larval and adult stages, respectively, is fundamentally different [171]. Accordingly, genetic variants affecting the expression of different genes have been identified to drive evolution in trichome patterns at different developmental stages [172,173]. Highly pleiotropic genes are important for different processes and thus they are involved in many gene regulatory networks. Therefore, the level of pleiotropy of a gene may also have an impact on its evolvability. Indeed, the expression of genes that are expressed in many tissues, i.e., pleiotropic factors, is more often conserved across species compared to tissue specific genes [174,175]. In summary, the architecture of gene regulatory networks influences how gene expression can evolve. A detailed understanding of tissue or stage specific gene regulatory networks and the integration of that knowledge into studies on the evolution of gene expression will certainly allow identifying general mechanisms generating variation in regulation.

6. Context Dependency Should be Considered in Comparative Expression Studies

In the light of context dependent gene expression and gene regulation it is important to account for these aspects in comparative expression studies. Since high throughput sequencing methods are highly sensitive, not all identified genes in such studies may be directly associated with the trait of interest, but rather represent background noise. Few studies specifically tested whether complex tissue composition influences the sensitivity to detect gene expression differences. A RNAseq experiment in *D. melanogaster* compared genome wide gene expression in central nervous system tissue between wildtype and transgenic flies after RNA interference (RNAi) mediated cell-type specific downregulation of a ubiquitously expressed gene. Intriguingly, the authors could show that contamination by surrounding tissue was sufficient to hamper the identification of the artificially downregulated gene [176]. This specific example strongly suggests that restricting sequencing efforts to the tissue and time point of interest allows identifying differentially expressed genes as specific as possible. The same rationale applies for eQTL and ASE studies, since the analysis of complex organs composed of various cell types does not allow gaining cell type specific mechanistic insights (Table 2).

The lack of tissue specificity can partially be accounted for by cell type or tissue enrichment in model organisms that offer a versatile transgenic toolkit (see also Table 2). This approach usually requires

the generation of transgenic individuals in which the target cell type or tissue is labeled by artificial fluorescent markers such as green fluorescent protein (GFP). Upon tissue dissociation, the labeled cells can be sorted by fluorescence-activated cell sorting (FACS) and classical bulk-RNAseq can be performed subsequently. This method has been successfully used to identify cell-type specific gene expression profiles [177–179] as well as to reveal candidate genes in evolutionary studies [180]. While this approach is restricted to genetically tractable model systems and requires in-depth information about the tissue of interest, recent advances in single cell RNA sequencing (scRNAseq) provide an excellent opportunity to gain cell type specific insights into gene expression of heterogeneous tissues without prior knowledge [124–126] (Table 2). A huge body of work has been published reporting for instance single cell atlases for various organisms such as embryos of *Drosophila melanogaster* [181], the cnidarian *Nematostella vectensis* [182], the planarian *Schmidtea mediterranea* [183] or the marine annelid *Platynereis dumerilii* [184]. Also, organ specific single cell atlases are being generated these days: In *Drosophila* for instance, new biological insights into the cell diversity, cell specific gene expression and gene regulation have been gained for entire aging brains [185], but also for parts of the brain such as the optic lobes [186] and the mid-brain [187]. The ability to reconstruct cell specific gene regulatory networks [188,189] from scRNAseq data provides the basis to relate comparative gene expression data to gene regulatory network architecture in a highly specific cellular context. Eventually, many high throughput sequencing applications to assess chromatin accessibility, histone modifications or transcription factor binding are applicable on single cell resolution (see Table 1 for details and references). Therefore, it is possible to compare the impact of genetic variation on these regulatory traits on single cell resolution. In summary, recent advances in single cell sequencing technologies provide an excellent opportunity to study context dependent gene expression in complex tissues. The combination of such methods with eQTL or ASE studies will allow revealing the impact of genetic variants on context dependent gene regulation.

Table 2. Comparison of different RNA sequencing methods.

	bulk-RNAseq of Whole Individuals	bulk-RNAseq with Prior Selection	scRNAseq
What can I do?			
Gain cell type specific gene expression	–	+/-	+
Identify overall gene expression profile	+	–	–
What do I need?			
Prior knowledge about the tissue or cells of interest	–	+	–
Transgenic organisms/fluorescently labeled cells	–	+	–
Specific technique to obtain tissue/cells	–	+/-	+

Even if a tissue or stage of interest was selected as specific as possible, candidate gene lists obtained by comparative expression studies can further be specified by integration of prior knowledge about the molecular functions of differentially expressed genes. While many genes show context dependent expression, housekeeping genes, which fulfil generic tasks in each cell, are often stably expressed across different tissues. Comparative approaches can be used to exclude generic differentially expressed genes by analyzing which transcripts are consistently differentially expressed across different tissues or time points and can therefore be removed from candidate gene lists (Figure 2). It is also helpful to have some prior knowledge about molecular pathways and processes that are involved in regulating the phenotypic trait of interest. Variation in physiological traits may be associated with hormonal signals

or enzymatic reactions, while morphological divergence is often linked to differences in underlying developmental processes. The growing Gene Ontology (GO) database coordinated by the Gene Ontology Consortium [190,191] provides an excellent basis for integrating differential gene expression and molecular functions. This tool allows to structure and categorize a list of candidate genes if no prior molecular or cellular knowledge for the trait of interest is available, by testing, whether a list of candidates is enriched in GO terms with a particular molecular or cellular function. Similarly, gene set enrichment analysis (GSEA) [192–195] can be employed to reveal if specific molecular or developmental pathways are involved in the development of the trait of interest [196]. Hence, the implementation of biological knowledge that recapitulates context specific information helps finding patterns in an otherwise unstructured dataset and helps to restrict the number of meaningful candidates.

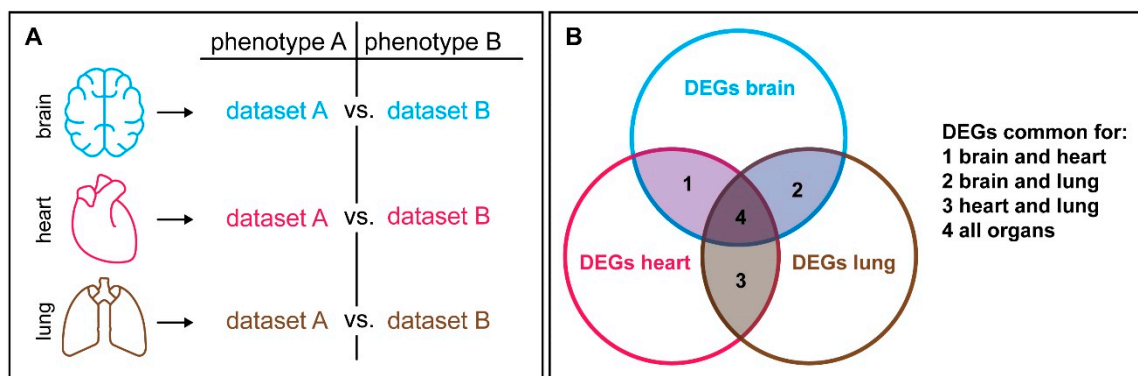


Figure 2. Generic factors that are expressed across different tissue can be excluded in correlation studies. (A) If specific candidate genes that are differentially expressed between phenotype A and B are supposed to be revealed, one can generate a comparable dataset for additional tissues. (B) Each pairwise comparison will reveal a certain number of differentially expressed genes (DEGs). The DEGs that are common in two (1–3) or all organs (4) are most likely generic factors that may be less informative for follow-up analyses.

7. Outlook

Although exciting insights into the evolution of gene expression and underlying regulatory mechanisms have been obtained, a few key questions remain to be answered.

It has for instance rarely been tested, whether intra- and interspecific differences in gene expression provide any fitness advantage for the organisms. Since gene expression divergence is often associated with phenotypic variation, one potential approach to answer this question could be the thorough integration of gene expression data with phenotypic data of relevant traits. Such a combinatorial approach has been successfully applied to reveal candidate genes involved in different nest building behaviors among the two mouse species *Peromyscus polionotus* and *P. maniculatus*. QTL mapping revealed 498 candidate genes associated with behavioral differences. Tissue specific RNA sequencing, accounting for the fact that the studied behavior is controlled in a specific brain region, identified 23 differentially expressed genes within the QTL region, of which nine genes showed signatures of *cis*-regulatory divergence. One of these nine genes was subsequently functionally validated [32]. Similar combinatorial approaches have been applied to identify key candidate genes responsible for variation in salt tolerance in rice (*Oryza rufipogon*) [197] and flowering time in rape (*Brassica napus*) [198]. Population genetics data that provides genome wide insights into signatures of selection can equally be combined with comparative expression data to reveal meaningful candidate genes underlying divergence of relevant phenotypic traits [199]. The integration of tissue and stage specific comparative expression data with quantitative and population genetics approaches thus provides a powerful way to reveal those differentially expressed genes with a direct effect on relevant phenotypic traits.

Another fundamental open question concerns the evolutionary forces underlying gene expression divergence. It will for instance be interesting to analyze the impact of genetic variants segregating

in natural populations of species on various levels of the gene regulation machinery. Since gene regulation is highly context dependent, it will be important to study causal links between genetic variants and gene expression and gene regulation in as many different organisms, tissues, cell types and developmental stages. With functional genomics and transcriptomics methods being widely applicable (Table 1), we can now test, whether mechanistic insights obtained by highly coordinated consortia studying individual human cell lines as well as tractable genetic model systems such as yeast and *Drosophila* hold true in other study systems.

8. Conclusions

Comparative genome-wide expression studies have been extensively used to reveal candidate factors to inform about the genotype–phenotype map (correlation studies) as well as to gain mechanistic insights into the evolution of gene regulation (eQTL and ASE). We argue that much more defined datasets must be generated in the future to fully account for the complexity and context dependency of gene regulation to increase the power to detect more meaningful candidate genes in correlation studies. We strongly believe that our current understanding of the evolution of gene expression provides a solid basis to incorporate new aspects of gene regulation, that are being revealed on a regular basis, to gain exciting new mechanistic insights into the evolutionary processes. There is still a sphere of cloudiness around the evolution of gene expression but digging deeper holds a chance of insight.

Author Contributions: Conceptualization, all authors; writing, review, and editing, all authors; visualization, N.P.; supervision, N.P.

Funding: Our work is funded by the Emmy Noether Programme of the Deutsche Forschungsgemeinschaft (grant number: PO 1648/3-1) to N.P.

Acknowledgments: We are grateful for the constructive suggestions by the two anonymous reviewers and the editors, which clearly improved the manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the writing of the manuscript.

References

1. Pearson, J.C.; Lemons, D.; McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* **2005**, *6*, 893–904. [[CrossRef](#)] [[PubMed](#)]
2. Hatefi, Y. The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem.* **1985**, *54*, 1015–1069. [[CrossRef](#)] [[PubMed](#)]
3. Saraste, M. Oxidative Phosphorylation at the fin de siècle. *Science* **1999**, *283*, 1488–1493. [[CrossRef](#)] [[PubMed](#)]
4. Janeway, C.A.; Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **2002**, *20*, 197–216. [[CrossRef](#)] [[PubMed](#)]
5. Hancock, R.E.W.; Haney, E.F.; Gill, E.E. The immunology of host defence peptides: Beyond antimicrobial activity. *Nat. Rev. Immunol.* **2016**, *16*, 321–334. [[CrossRef](#)]
6. Lumpkin, E.A.; Caterina, M.J. Mechanisms of sensory transduction in the skin. *Nature* **2007**, *445*, 858–865. [[CrossRef](#)] [[PubMed](#)]
7. Yau, K.-W.; Hardie, R.C. Phototransduction motifs and variations. *Cell* **2009**, *139*, 246–264. [[CrossRef](#)]
8. Avery, O.T. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J. Exp. Med.* **1944**, *79*, 137–158. [[CrossRef](#)]
9. Thoday, J.M. Location of Polygenes. *Nat. Insight Biodivers.* **1961**, *191*, 368–370. [[CrossRef](#)]
10. Lander, E.S.; Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **1989**, *121*, 185–199.
11. Roff, D.A. A centennial celebration for quantitative genetics. *Evol. Int. J. Org. Evol.* **2007**, *61*, 1017–1032. [[CrossRef](#)] [[PubMed](#)]

12. Mackay, T.F.C.; Richards, S.; Stone, E.a.; Barbadilla, A.; Ayroles, J.F.; Zhu, D.; Casillas, S.; Han, Y.; Magwire, M.M.; Cridland, J.M.; et al. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **2012**, *482*, 173–178. [[CrossRef](#)] [[PubMed](#)]
13. Linnen, C.R.; Poh, Y.-P.; Peterson, B.K.; Barrett, R.D.H.; Larson, J.G.; Jensen, J.D.; Hoekstra, H.E. Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. *Science* **2013**, *339*, 1312–1316. [[CrossRef](#)] [[PubMed](#)]
14. Hoekstra, H.E.; Hirschmann, R.J.; Bunday, R.A.; Insel, P.A.; Crossland, J.P. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **2006**, *313*, 101–104. [[CrossRef](#)] [[PubMed](#)]
15. Lang, M.; Murat, S.; Clark, A.G.; Gouppil, G.; Blais, C.; Matzkin, L.M.; Guittard, E.; Yoshiyama-Yanagawa, T.; Kataoka, H.; Niwa, R.; et al. Mutations in the neverland Gene Turned *Drosophila pachea* into an Obligate Specialist Species. *Science* **2012**, *337*, 1658–1661. [[CrossRef](#)] [[PubMed](#)]
16. Weinberger, S.; Topping, M.P.; Yan, J.; Claeys, A.; Geest, N.D.; Ozbay, D.; Hassan, T.; He, X.; Albert, J.T.; Hassan, B.A.; et al. Evolutionary changes in transcription factor coding sequence quantitatively alter sensory organ development and function. *eLife* **2017**, *6*. [[CrossRef](#)] [[PubMed](#)]
17. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **2009**, *10*, 57–63. [[CrossRef](#)] [[PubMed](#)]
18. Oppenheim, S.J.; Baker, R.H.; Simon, S.; DeSalle, R. We can't all be supermodels: The value of comparative transcriptomics to the study of non-model insects. *Insect Mol. Biol.* **2015**, *24*, 139–154. [[CrossRef](#)]
19. Necsulea, A.; Kaessmann, H. Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* **2014**, *15*, 734–748. [[CrossRef](#)]
20. Alvarez, M.; Schrey, A.W.; Richards, C.L. Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Mol. Ecol.* **2015**, *24*, 710–725. [[CrossRef](#)]
21. Todd, E.V.; Black, M.A.; Gemmell, N.J. The power and promise of RNA-seq in ecology and evolution. *Mol. Ecol.* **2016**, *25*, 1224–1241. [[CrossRef](#)] [[PubMed](#)]
22. Arif, S.; Kittelmann, S.; McGregor, A.P. From shavenbaby to the naked valley: Trichome formation as a model for evolutionary developmental biology. *Evol. Dev.* **2015**, *17*, 120–126. [[CrossRef](#)] [[PubMed](#)]
23. Preger-Ben Noon, E.; Sabarís, G.; Ortiz, D.M.; Sager, J.; Liebowitz, A.; Stern, D.L.; Frankel, N. Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function. *Cell Rep.* **2018**, *22*, 3021–3031. [[CrossRef](#)] [[PubMed](#)]
24. Rebeiz, M.; POOL, J.E.; Kassner, V.A.; AQUADRO, C.F.; Carroll, S.B. Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* **2009**, *326*, 1663–1667. [[CrossRef](#)] [[PubMed](#)]
25. Gompel, N.; Prud'homme, B.; Wittkopp, P.J.; Kassner, V.a.; Carroll, S.B. Chance caught on the wing: Cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **2005**, *433*, 481–487. [[CrossRef](#)] [[PubMed](#)]
26. Manceau, M.; Domingues, V.S.; Mallarino, R.; Hoekstra, H.E. The developmental role of Agouti in color pattern evolution. *Science* **2011**, *331*, 1062–1065. [[CrossRef](#)]
27. Hoekstra, H.E. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **2006**, *97*, 222–234. [[CrossRef](#)]
28. Kratochwil, C.F.; Liang, Y.; Gerwin, J.; Woltering, J.M.; Urban, S.; Henning, F.; Machado-Schiaffino, G.; Hulsey, C.D.; Meyer, A. Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations. *Science* **2018**, *362*, 457–460. [[CrossRef](#)]
29. Hagen, J.F.D.; Mendes, C.C.; Tanaka, K.M.; Gaspar, P.; Kittelmann, M.; McGregor, A.P.; Nunes, M.D.S. tartan underlies the evolution of male genital morphology. *bioRxiv* **2018**. [[CrossRef](#)]
30. Nagy, O.; Nuez, I.; Savisaar, R.; Peluffo, A.E.; Yassin, A.; Lang, M.; Stern, D.L.; Matute, D.R.; David, J.R.; Courtier-Orgogozo, V. Correlated Evolution of Two Copulatory Organs via a Single cis-Regulatory Nucleotide Change. *Curr. Biol.* **2018**, *28*, 3450–3457. [[CrossRef](#)]
31. Kocher, S.D.; Mallarino, R.; Rubin, B.E.R.; Yu, D.W.; Hoekstra, H.E.; Pierce, N.E. The genetic basis of a social polymorphism in halictid bees. *Nat. Commun.* **2018**, *9*, 4338. [[CrossRef](#)] [[PubMed](#)]
32. Bendesky, A.; Kwon, Y.-M.; Lassance, J.-M.; Lewarch, C.L.; Yao, S.; Peterson, B.K.; He, M.X.; Dulac, C.; Hoekstra, H.E. The genetic basis of parental care evolution in monogamous mice. *Nature* **2017**, *544*, 434–439. [[CrossRef](#)] [[PubMed](#)]

33. Huang, W.; Richards, S.; Carbone, M.A.; Zhu, D.; Anholt, R.R.H.; Ayroles, J.F.; Duncan, L.; Jordan, K.W.; Lawrence, F.; Magwire, M.M.; et al. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 15553–15559. [[CrossRef](#)] [[PubMed](#)]
34. Zhou, S.; Luoma, S.E.; St Armour, G.E.; Thakkar, E.; Mackay, T.F.C.; Anholt, R.R.H. A *Drosophila* model for toxicogenomics: Genetic variation in susceptibility to heavy metal exposure. *PLoS Genet.* **2017**, *13*, e1006907. [[CrossRef](#)] [[PubMed](#)]
35. Wang, J.; Ding, J.; Tan, B.; Robinson, K.M.; Michelson, I.H.; Johansson, A.; Nystedt, B.; Scofield, D.G.; Nilsson, O.; Jansson, S.; et al. A major locus controls local adaptation and adaptive life history variation in a perennial plant. *Genome Biol.* **2018**, *19*, 72. [[CrossRef](#)] [[PubMed](#)]
36. Chang, J.; Zhou, Y.; Hu, X.; Lam, L.; Henry, C.; Green, E.M.; Kita, R.; Kobor, M.S.; Fraser, H.B. The molecular mechanism of a cis-regulatory adaptation in yeast. *PLoS Genet.* **2013**, *9*, e1003813. [[CrossRef](#)]
37. Catalán, A.; Glaser-Schmitt, A.; Argyridou, E.; Duchen, P.; Parsch, J. An Indel Polymorphism in the MtnA 3' Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in *Drosophila melanogaster*. *PLoS Genet.* **2016**, *12*, e1005987. [[CrossRef](#)]
38. Hutter, S.; Saminadin-Peter, S.S.; Stephan, W.; Parsch, J. Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biol.* **2008**, *9*, R12. [[CrossRef](#)]
39. Townsend, J.P.; Cavalieri, D.; Hartl, D.L. Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* **2003**, *20*, 955–963. [[CrossRef](#)]
40. Oleksiak, M.F.; Churchill, G.A.; Crawford, D.L. Variation in gene expression within and among natural populations. *Nat. Genet.* **2002**, *32*, 261–266. [[CrossRef](#)]
41. Signor, S.A.; Nuzhdin, S.V. The Evolution of Gene Expression in cis and trans. *Trends Genet.* **2018**, *34*, 532–544. [[CrossRef](#)] [[PubMed](#)]
42. Nourmohammad, A.; Rambeau, J.; Held, T.; Kovacova, V.; Berg, J.; Lässig, M. Adaptive Evolution of Gene Expression in *Drosophila*. *Cell Rep.* **2017**, *20*, 1385–1395. [[CrossRef](#)] [[PubMed](#)]
43. Zhang, Y.; Sturgill, D.; Parisi, M.; Kumar, S.; Oliver, B. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* **2007**, *450*, 233–237. [[CrossRef](#)] [[PubMed](#)]
44. Guschanski, K.; Warnefors, M.; Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **2017**, *27*, 1461–1474. [[CrossRef](#)] [[PubMed](#)]
45. Rockman, M.V.; Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **2006**, *7*, 862–872. [[CrossRef](#)]
46. Gilad, Y.; Rifkin, S.A.; Pritchard, J.K. Revealing the architecture of gene regulation: The promise of eQTL studies. *Trends Genet.* **2008**, *24*, 408–415. [[CrossRef](#)]
47. Jia, Z.; Xu, S. Mapping quantitative trait loci for expression abundance. *Genetics* **2007**, *176*, 611–623. [[CrossRef](#)]
48. Dixon, A.L.; Liang, L.; Moffatt, M.F.; Chen, W.; Heath, S.; Wong, K.C.C.; Taylor, J.; Burnett, E.; Gut, I.; Farrall, M.; et al. A genome-wide association study of global gene expression. *Nat. Genet.* **2007**, *39*, 1202–1207. [[CrossRef](#)]
49. Knight, J.C. Allele-specific gene expression uncovered. *Trends Genet.* **2004**, *20*, 113–116. [[CrossRef](#)]
50. Coolon, J.D.; Wittkopp, P.J. cis- and trans -Regulation in *Drosophila* Interspecific Hybrids. *Polyploid Hybrid Genomics* **2015**, *305*, 37–57. [[CrossRef](#)]
51. Korir, P.K.; Seoighe, C. Inference of allele-specific expression from RNA-seq data. *Methods Mol. Biol.* **2014**, *1112*, 49–69. [[CrossRef](#)] [[PubMed](#)]
52. Osada, N.; Miyagi, R.; Takahashi, A. Cis- and Trans-regulatory Effects on Gene Expression in a Natural Population of *Drosophila melanogaster*. *Genetics* **2017**, *206*, 2139–2148. [[CrossRef](#)] [[PubMed](#)]
53. Graze, R.M.; McIntyre, L.M.; Main, B.J.; Wayne, M.L.; Nuzhdin, S.V. Regulatory divergence in *Drosophila melanogaster* and *D. simulans*, a genomewide analysis of allele-specific expression. *Genetics* **2009**, *183*, 547–561. [[CrossRef](#)] [[PubMed](#)]
54. Graze, R.M.; Novelo, L.L.; Amin, V.; Fear, J.M.; Casella, G.; Nuzhdin, S.V.; McIntyre, L.M. Allelic imbalance in *Drosophila* hybrid heads: Exons, isoforms, and evolution. *Mol. Biol. Evol.* **2012**, *29*, 1521–1532. [[CrossRef](#)] [[PubMed](#)]
55. Fontanillas, P.; Landry, C.R.; Wittkopp, P.J.; Russ, C.; Gruber, J.D.; Nusbaum, C.; Hartl, D.L. Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Mol. Ecol.* **2010**, *19*, 212–227. [[CrossRef](#)] [[PubMed](#)]
56. McManus, C.J.; Coolon, J.D.; Duff, M.O.; Eipper-Mains, J.; Graveley, B.R.; Wittkopp, P.J. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* **2010**, *20*, 816–825. [[CrossRef](#)] [[PubMed](#)]

57. Tirosh, I.; Reikhav, S.; Levy, A.a.; Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **2009**, *324*, 659–662. [[CrossRef](#)]
58. Zhang, X.; Borevitz, J.O. Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* **2009**, *182*, 943–954. [[CrossRef](#)]
59. Lemmon, Z.H.; Bukowski, R.; Sun, Q.; Doebley, J.F. The Role of cis Regulatory Evolution in Maize Domestication. *PLoS Genet.* **2014**, *10*. [[CrossRef](#)]
60. Lawniczak, M.K.N.; Holloway, A.K.; Begun, D.J.; Jones, C.D. Genomic analysis of the relationship between gene expression variation and DNA polymorphism in *Drosophila simulans*. *Genome Biol.* **2008**, *9*, R125. [[CrossRef](#)]
61. Lappalainen, T.; Sammeth, M.; Friedländer, M.R.; AC't Hoen, P.; Monlong, J.; Rivas, M.A.; González-Porta, M.; Kurbatova, N.; Griebel, T.; Ferreira, P.G.; et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **2013**, *501*, 506–511. [[CrossRef](#)] [[PubMed](#)]
62. Lübbe, A.; Schaffner, W. Tissue-specific gene expression. *Trends Neurosci.* **1985**, *8*, 100–104. [[CrossRef](#)]
63. 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. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)] [[PubMed](#)]
64. Sexton, T.; Yaffe, E.; Kenigsberg, E.; Bantignies, F.; Leblanc, B.; Hoichman, M.; Parrinello, H.; Tanay, A.; Cavalli, G. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **2012**, *148*, 458–472. [[CrossRef](#)] [[PubMed](#)]
65. Nora, E.P.; Lajoie, B.R.; Schulz, E.G.; Giorgetti, L.; Okamoto, I.; Servant, N.; Piolot, T.; van Berkum, N.L.; Meisig, J.; Sedat, J.; et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **2012**, *485*, 381–385. [[CrossRef](#)]
66. Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J.S.; Ren, B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380. [[CrossRef](#)] [[PubMed](#)]
67. Andrey, G.; Montavon, T.; Mascrez, B.; Gonzalez, F.; Noordermeer, D.; Leleu, M.; Trono, D.; Spitz, F.; Duboule, D. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **2013**, *340*, 1234167. [[CrossRef](#)] [[PubMed](#)]
68. Javierre, B.M.; Burren, O.S.; Wilder, S.P.; Kreuzhuber, R.; Hill, S.M.; Sewitz, S.; Cairns, J.; Wingett, S.W.; Várnai, C.; Thiecke, M.J.; et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **2016**, *167*, 1369–1384. [[CrossRef](#)]
69. Buenostro, J.D.; Giresi, P.G.; Zaba, L.C.; Chang, H.Y.; Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **2013**, *10*, 1213–1218. [[CrossRef](#)]
70. Cusanovich, D.A.; Hill, A.J.; Aghamirzaie, D.; Daza, R.M.; Pliner, H.A.; Berletch, J.B.; Filippova, G.N.; Huang, X.; Christiansen, L.; DeWitt, W.S.; et al. A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **2018**. [[CrossRef](#)]
71. Cusanovich, D.A.; Reddington, J.P.; Garfield, D.A.; Daza, R.M.; Aghamirzaie, D.; Marco-Ferreres, R.; Pliner, H.A.; Christiansen, L.; Qiu, X.; Steemers, F.J.; et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **2018**, *555*, 538–542. [[CrossRef](#)] [[PubMed](#)]
72. Ho, J.W.K.; Jung, Y.L.; Liu, T.; Alver, B.H.; Lee, S.; Ikegami, K.; Sohn, K.-A.; Minoda, A.; Tolstorukov, M.Y.; Appert, A.; et al. Comparative analysis of metazoan chromatin organization. *Nature* **2014**, *512*, 449–452. [[CrossRef](#)] [[PubMed](#)]
73. Liu, T.; Rechtsteiner, A.; Egelhofer, T.A.; Vielle, A.; Latorre, I.; Cheung, M.-S.; Ercan, S.; Ikegami, K.; Jensen, M.; Kolasinska-Zwierz, P.; et al. Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res.* **2011**, *21*, 227–236. [[CrossRef](#)] [[PubMed](#)]
74. Lawrence, M.; Daujat, S.; Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **2016**, *32*, 42–56. [[CrossRef](#)] [[PubMed](#)]
75. Gates, L.A.; Foulds, C.E.; O'Malley, B.W. Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends Biochem. Sci.* **2017**, *42*, 977–989. [[CrossRef](#)] [[PubMed](#)]
76. Bird, A.P.; Wolffe, A.P. Methylation-Induced Repression—Belts, Braces, and Chromatin. *Cell* **1999**, *99*, 451–454. [[CrossRef](#)]

77. Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nat. Genet.* **2003**, *33*, 245–254. [[CrossRef](#)] [[PubMed](#)]
78. Zhu, H.; Wang, G.; Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* **2016**, *17*, 551–565. [[CrossRef](#)]
79. Mohn, F.; Schübeler, D. Genetics and epigenetics: Stability and plasticity during cellular differentiation. *Trends Genet.* **2009**, *25*, 129–136. [[CrossRef](#)]
80. Engreitz, J.M.; Pandya-Jones, A.; McDonel, P.; Shishkin, A.; Sirokman, K.; Surka, C.; Kadri, S.; Xing, J.; Goren, A.; Lander, E.S.; et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **2013**, *341*, 1237973. [[CrossRef](#)]
81. Pandey, R.R.; Mondal, T.; Mohammad, F.; Enroth, S.; Redrup, L.; Komorowski, J.; Nagano, T.; Mancini-Dinardo, D.; Kanduri, C. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* **2008**, *32*, 232–246. [[CrossRef](#)] [[PubMed](#)]
82. Wang, K.C.; Yang, Y.W.; Liu, B.; Sanyal, A.; Corces-Zimmerman, R.; Chen, Y.; Lajoie, B.R.; Protacio, A.; Flynn, R.A.; Gupta, R.A.; et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **2011**, *472*, 120–124. [[CrossRef](#)] [[PubMed](#)]
83. Li, W.; Notani, D.; Ma, Q.; Tanasa, B.; Nunez, E.; Chen, A.Y.; Merkurjev, D.; Zhang, J.; Ohgi, K.; Song, X.; et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **2013**, *498*, 516–520. [[CrossRef](#)] [[PubMed](#)]
84. Xiao, M.; Li, J.; Li, W.; Wang, Y.; Wu, F.; Xi, Y.; Zhang, L.; Ding, C.; Luo, H.; Li, Y.; et al. MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol.* **2017**, *14*, 1326–1334. [[CrossRef](#)] [[PubMed](#)]
85. Berezikov, E.; Robine, N.; Samsonova, A.; Westholm, J.O.; Naqvi, A.; Hung, J.-H.; Okamura, K.; Dai, Q.; Bortolamiol-Becet, D.; Martin, R.; et al. Deep annotation of *Drosophila melanogaster* microRNAs yields insights into their processing, modification, and emergence. *Genome Res.* **2011**, *21*, 203–215. [[CrossRef](#)] [[PubMed](#)]
86. Mohammed, J.; Flynt, A.S.; Panzarino, A.M.; Mondal, M.M.H.; DeCruz, M.; Siepel, A.; Lai, E.C. Deep experimental profiling of microRNA diversity, deployment, and evolution across the *Drosophila* genus. *Genome Res.* **2018**, *28*, 52–65. [[CrossRef](#)] [[PubMed](#)]
87. Ludwig, N.; Leidinger, P.; Becker, K.; Backes, C.; Fehlmann, T.; Pallasch, C.; Rheinheimer, S.; Meder, B.; Stähler, C.; Meese, E.; et al. Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* **2016**, *44*, 3865–3877. [[CrossRef](#)]
88. Spitz, F.; Furlong, E.E.M. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **2012**, *13*, 613–626. [[CrossRef](#)]
89. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [[CrossRef](#)]
90. Homem, C.C.F.; Knoblich, J.A. *Drosophila* neuroblasts: A model for stem cell biology. *Development* **2012**, *139*, 4297–4310. [[CrossRef](#)]
91. Karcavich, R.E. Generating neuronal diversity in the *Drosophila* central nervous system: A view from the ganglion mother cells. *Dev. Dyn.* **2005**, *232*, 609–616. [[CrossRef](#)] [[PubMed](#)]
92. Technau, G.M.; Berger, C.; Urbach, R. Generation of cell diversity and segmental pattern in the embryonic central nervous system of *Drosophila*. *Dev. Dyn.* **2006**, *235*, 861–869. [[CrossRef](#)] [[PubMed](#)]
93. Nogales, E.; Louder, R.K.; He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu. Rev. Biophys.* **2017**, *46*, 59–83. [[CrossRef](#)] [[PubMed](#)]
94. Engel, C.; Neyer, S.; Cramer, P. Distinct Mechanisms of Transcription Initiation by RNA Polymerases I and II. *Annu. Rev. Biophys.* **2018**, *47*, 425–446. [[CrossRef](#)] [[PubMed](#)]
95. Levine, M.; Tjian, R. Transcription regulation and animal diversity. *Nature* **2003**, *424*, 147–151. [[CrossRef](#)] [[PubMed](#)]
96. Banerji, J.; Rusconi, S.; Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **1981**, *27*, 299–308. [[CrossRef](#)]
97. Kirchhamer, C.V.; Yuh, C.H.; Davidson, E.H. Modular cis-regulatory organization of developmentally expressed genes: Two genes transcribed territorially in the sea urchin embryo, and additional examples. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9322–9328. [[CrossRef](#)]

98. Arnone, M.I.; Davidson, E.H. The hardwiring of development: Organization and function of genomic regulatory systems. *Development* **1997**, *124*, 1851–1864.
99. Small, S.; Blair, A.; Levine, M. Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* **1992**, *11*, 4047–4057. [[CrossRef](#)]
100. Goto, T.; Macdonald, P.; Maniatis, T. Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **1989**, *57*, 413–422. [[CrossRef](#)]
101. Haenlin, M.; Cubadda, Y.; Blondeau, F.; Heitzler, P.; Lutz, Y.; Simpson, P.; Romain, P. Transcriptional activity of Pannier is regulated negatively by heterodimerization of the GATA DNA-binding domain with a cofactor encoded by the u-shaped gene of *Drosophila*. *Genes Dev.* **1997**, *11*, 3096–3108. [[CrossRef](#)] [[PubMed](#)]
102. Fromental-Romain, C.; Taquet, N.; Romain, P. Transcriptional interactions between the pannier isoforms and the cofactor U-shaped during neural development in *Drosophila*. *Mech. Dev.* **2010**, *127*, 442–457. [[CrossRef](#)] [[PubMed](#)]
103. Fromental-Romain, C.; Vanolst, L.; Delaporte, C.; Romain, P. pannier encodes two structurally related isoforms that are differentially expressed during *Drosophila* development and display distinct functions during thorax patterning. *Mech. Dev.* **2008**, *125*, 43–57. [[CrossRef](#)] [[PubMed](#)]
104. Meireles-Filho, A.C.a.; Bardet, A.F.; Yáñez-Cuna, J.O.; Stampfel, G.; Stark, A. cis-regulatory requirements for tissue-specific programs of the circadian clock. *Curr. Biol.* **2014**, *24*, 1–10. [[CrossRef](#)] [[PubMed](#)]
105. Elkon, R.; Ugalde, A.P.; Agami, R. Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* **2013**, *14*, 496–506. [[CrossRef](#)] [[PubMed](#)]
106. Mayr, C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol.* **2016**, *26*, 227–237. [[CrossRef](#)]
107. Topisirovic, I.; Svitkin, Y.V.; Sonenberg, N.; Shatkin, A.J. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 277–298. [[CrossRef](#)]
108. Matera, A.G.; Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 108–121. [[CrossRef](#)]
109. Grosso, A.R.; Gomes, A.Q.; Barbosa-Morais, N.L.; Caldeira, S.; Thorne, N.P.; Grech, G.; von Lindern, M.; Carmo-Fonseca, M. Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res.* **2008**, *36*, 4823–4832. [[CrossRef](#)]
110. Bartel, D.P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **2004**, *116*, 281–297. [[CrossRef](#)]
111. Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21. [[CrossRef](#)] [[PubMed](#)]
112. Gong, C.; Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **2011**, *470*, 284–288. [[CrossRef](#)] [[PubMed](#)]
113. Kretz, M.; Siprashvili, Z.; Chu, C.; Webster, D.E.; Zehnder, A.; Qu, K.; Lee, C.S.; Flockhart, R.J.; Groff, A.F.; Chow, J.; et al. Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature* **2013**, *493*, 231–235. [[CrossRef](#)] [[PubMed](#)]
114. Poliseno, L.; Salmena, L.; Zhang, J.; Carver, B.; Haveman, W.J.; Pandolfi, P.P. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **2010**, *465*, 1033–1038. [[CrossRef](#)]
115. Alvarez-Garcia, I.; Miska, E.A. MicroRNA functions in animal development and human disease. *Development* **2005**, *132*, 4653–4662. [[CrossRef](#)] [[PubMed](#)]
116. Agirre, X.; Meydan, C.; Jiang, Y.; Garate, L.; Doane, A.S.; Li, Z.; Verma, A.; Paiva, B.; Martín-Subero, J.I.; Elemento, O.; et al. Long non-coding RNAs discriminate the stages and gene regulatory states of human humoral immune response. *Nat. Commun.* **2019**, *10*, 821. [[CrossRef](#)]
117. Schor, I.E.; Bussotti, G.; Maleš, M.; Forneris, M.; Viales, R.R.; Enright, A.J.; Furlong, E.E.M. Non-coding RNA Expression, Function, and Variation during *Drosophila* Embryogenesis. *Curr. Biol.* **2018**, *28*, 3547–3561. [[CrossRef](#)]
118. Mercer, T.R.; Mattick, J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300–307. [[CrossRef](#)]
119. O'Neil, D.; Glowatz, H.; Schlumpberger, M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr. Protoc. Mol. Biol.* **2013**, 1–8. [[CrossRef](#)]
120. Coenen-Stass, A.M.L.; Magen, I.; Brooks, T.; Ben-Dov, I.Z.; Greensmith, L.; Hornstein, E.; Fratta, P. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol.* **2018**, *15*, 1133–1145. [[CrossRef](#)]

121. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [[CrossRef](#)] [[PubMed](#)]
122. Pritchard, C.C.; Cheng, H.H.; Tewari, M. MicroRNA profiling: Approaches and considerations. *Nat. Rev. Genet.* **2012**, *13*, 358. [[CrossRef](#)] [[PubMed](#)]
123. Akhade, V.S.; Pal, D.; Kanduri, C. Long Noncoding RNA: Genome Organization and Mechanism of Action. *Adv. Exp. Med. Biol.* **2017**, *1008*, 47–74. [[CrossRef](#)]
124. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A.; et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Meth.* **2009**, *6*, 377–382. [[CrossRef](#)] [[PubMed](#)]
125. Zheng, G.X.Y.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **2017**, *8*, 14049. [[CrossRef](#)] [[PubMed](#)]
126. Svensson, V.; Vento-Tormo, R.; Teichmann, S.A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **2018**, *13*, 599–604. [[CrossRef](#)] [[PubMed](#)]
127. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **2015**, *2015*, 21–29. [[CrossRef](#)]
128. Corces, M.R.; Trevino, A.E.; Hamilton, E.G.; Greenside, P.G.; Sinnott-Armstrong, N.A.; Vesuna, S.; Satpathy, A.T.; Rubin, A.J.; Montine, K.S.; Wu, B.; et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **2017**, *14*, 959–962. [[CrossRef](#)]
129. Daugherty, A.C.; Yeo, R.W.; Buenrostro, J.D.; Greenleaf, W.J.; Kundaje, A.; Brunet, A. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Res.* **2017**, *27*, 2096–2107. [[CrossRef](#)]
130. Cusanovich, D.A.; Daza, R.; Adey, A.; Pliner, H.A.; Christiansen, L.; Gunderson, K.L.; Steemers, F.J.; Trapnell, C.; Shendure, J. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **2015**, *348*, 910–914. [[CrossRef](#)]
131. Buenrostro, J.D.; Wu, B.; Litzenburger, U.M.; Ruff, D.; Gonzales, M.L.; Snyder, M.P.; Chang, H.Y.; Greenleaf, W.J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nat. Insight Biodivers.* **2015**, *523*, 486–490. [[CrossRef](#)] [[PubMed](#)]
132. Johnson, D.S.; Mortazavi, A.; Myers, R.M.; Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **2007**, *316*, 1497–1502. [[CrossRef](#)] [[PubMed](#)]
133. Robertson, G.; Hirst, M.; Bainbridge, M.; Bilenky, M.; Zhao, Y.; Zeng, T.; Euskirchen, G.; Bernier, B.; Varhol, R.; Delaney, A.; et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **2007**, *4*, 651–657. [[CrossRef](#)] [[PubMed](#)]
134. Schmidt, D.; Wilson, M.D.; Ballester, B.; Schwalie, P.C.; Brown, G.D.; Marshall, A.; Kutter, C.; Watt, S.; Martinez-Jimenez, C.P.; Mackay, S.; et al. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **2010**, *328*, 1036–1040. [[CrossRef](#)] [[PubMed](#)]
135. Rotem, A.; Ram, O.; Shores, N.; Sperling, R.A.; Goren, A.; Weitz, D.A.; Bernstein, B.E. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **2015**, *33*, 1165–1172. [[CrossRef](#)] [[PubMed](#)]
136. Belton, J.-M.; McCord, R.P.; Gibcus, J.H.; Naumova, N.; Zhan, Y.; Dekker, J. Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **2012**, *58*, 268–276. [[CrossRef](#)]
137. Sati, S.; Cavalli, G. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* **2017**, *126*, 33–44. [[CrossRef](#)]
138. Ma, W.; Ay, F.; Lee, C.; Gulsoy, G.; Deng, X.; Cook, S.; Hesson, J.; Cavanaugh, C.; Ware, C.B.; Krumm, A.; et al. Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat. Methods* **2015**, *12*, 71–78. [[CrossRef](#)]
139. Hsieh, T.-H.S.; Weiner, A.; Lajoie, B.; Dekker, J.; Friedman, N.; Rando, O.J. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **2015**, *162*, 108–119. [[CrossRef](#)]
140. Fishman, V.; Battulin, N.; Nuriddinov, M.; Maslova, A.; Zlotina, A.; Strunov, A.; Chervyakova, D.; Korablev, A.; Serov, O.; Krasikova, A. 3D organization of chicken genome demonstrates evolutionary conservation of topologically associated domains and highlights unique architecture of erythrocytes' chromatin. *Nucleic Acids Res.* **2019**, *47*, 648–665. [[CrossRef](#)]

141. Nagano, T.; Lubling, Y.; Stevens, T.J.; Schoenfelder, S.; Yaffe, E.; Dean, W.; Laue, E.D.; Tanay, A.; Fraser, P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nat. Insight Biodivers.* **2013**, *502*, 59–64. [[CrossRef](#)]
142. Frommer, M.; McDonald, L.E.; Millar, D.S.; Collis, C.M.; Watt, F.; Grigg, G.W.; Molloy, P.L.; Paul, C.L. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1827–1831. [[CrossRef](#)]
143. Cokus, S.J.; Feng, S.; Zhang, X.; Chen, Z.; Merriman, B.; Haudenschild, C.D.; Pradhan, S.; Nelson, S.F.; Pellegrini, M.; Jacobsen, S.E. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nat. Insight Biodivers.* **2008**, *452*, 215–219. [[CrossRef](#)]
144. Lea, A.J.; Vilgalys, T.P.; Durst, P.A.P.; Tung, J. Maximizing ecological and evolutionary insight in bisulfite sequencing data sets HHS Public Access. *Nat. Ecol. Evol.* **2017**, *1*, 1074–1083. [[CrossRef](#)]
145. Yong, W.-S.; Hsu, F.-M.; Chen, P.-Y. Profiling genome-wide DNA methylation. *Epigenet. Chromatin* **2016**, *9*, 26. [[CrossRef](#)]
146. Verhoeven, K.J.F.; VonHoldt, B.M.; Sork, V.L. Epigenetics in ecology and evolution: What we know and what we need to know. *Mol. Ecol.* **2016**, *25*, 1631–1638. [[CrossRef](#)]
147. Guo, H.; Zhu, P.; Wu, X.; Li, X.; Wen, L.; Tang, F. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **2013**, *23*, 2126–2135. [[CrossRef](#)] [[PubMed](#)]
148. Smallwood, S.A.; Lee, H.J.; Angermueller, C.; Krueger, F.; Saadeh, H.; Peat, J.; Andrews, S.R.; Stegle, O.; Reik, W.; Kelsey, G. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **2014**, *11*, 817–820. [[CrossRef](#)] [[PubMed](#)]
149. Brawand, D.; Soumillon, M.; Necsulea, A.; Julien, P.; Csárdi, G.; Harrigan, P.; Weier, M.; Liechti, A.; Aximu-Petri, A.; Kircher, M.; et al. The evolution of gene expression levels in mammalian organs. *Nature* **2011**, *478*, 343–348. [[CrossRef](#)] [[PubMed](#)]
150. Irie, N.; Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* **2011**, *2*, 248. [[CrossRef](#)]
151. Domazet-Lošo, T.; Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **2010**, *468*, 815–818. [[CrossRef](#)] [[PubMed](#)]
152. Liscovitch, N.; Chechik, G. Specialization of gene expression during mouse brain development. *PLoS Comput. Biol.* **2013**, *9*, e1003185. [[CrossRef](#)] [[PubMed](#)]
153. Spies, N.; Smith, C.L.; Rodriguez, J.M.; Baker, J.C.; Batzoglou, S.; Sidow, A. Constraint and divergence of global gene expression in the mammalian embryo. *eLife* **2015**, *4*, e05538. [[CrossRef](#)] [[PubMed](#)]
154. Yang, X.; Schadt, E.E.; Wang, S.; Wang, H.; Arnold, A.P.; Ingram-Drake, L.; Drake, T.A.; Lusis, A.J. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **2006**, *16*, 995–1004. [[CrossRef](#)] [[PubMed](#)]
155. Wittkopp, P.J.; Haerum, B.K.; Clark, A.G. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **2008**, *40*, 346–350. [[CrossRef](#)]
156. Glaser-Schmitt, A.; Parsch, J. Functional characterization of adaptive variation within a cis-regulatory element influencing *Drosophila melanogaster* growth. *PLoS Biol.* **2018**, *16*, e2004538. [[CrossRef](#)] [[PubMed](#)]
157. Frazer, K.A.; Ballinger, D.G.; Cox, D.R.; Hinds, D.A.; Stuve, L.L.; Gibbs, R.A.; Belmont, J.W.; Boudreau, A.; Hardenbol, P.; Leal, S.M.; et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **2007**, *449*, 851–861. [[CrossRef](#)]
158. Grubert, F.; Zaugg, J.B.; Kasowski, M.; Ursu, O.; Spacek, D.V.; Martin, A.R.; Greenside, P.; Srivas, R.; Phanstiel, D.H.; Pekowska, A.; et al. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell* **2015**, *162*, 1051–1065. [[CrossRef](#)]
159. 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. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **2012**, *482*, 390–394. [[CrossRef](#)]
160. McVicker, G.; van de Geijn, B.; Degner, J.F.; Cain, C.E.; Banovich, N.E.; Raj, A.; Lewellen, N.; Myrthil, M.; Gilad, Y.; Pritchard, J.K. Identification of genetic variants that affect histone modifications in human cells. *Science* **2013**, *342*, 747–749. [[CrossRef](#)]

161. Del Rosario, R.C.-H.; Poschmann, J.; Rouam, S.L.; Png, E.; Khor, C.C.; Hibberd, M.L.; Prabhakar, S. Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat. Methods* **2015**, *12*, 458–464. [[CrossRef](#)] [[PubMed](#)]
162. Li, Y.L.; van de Geijn, B.; Raj, A.; Knowles, D.A.; Petti, A.A.; Golan, D.; Gilad, Y.; Pritchard, J.K. RNA splicing is a primary link between genetic variation and disease. *Science* **2016**, *352*, 600–604. [[CrossRef](#)] [[PubMed](#)]
163. Monlong, J.; Calvo, M.; Ferreira, P.G.; Guigó, R. Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat. Commun.* **2014**, *5*, 4698. [[CrossRef](#)] [[PubMed](#)]
164. Han, J.; Xiong, J.; Wang, D.; Fu, X.-D. Pre-mRNA splicing: Where and when in the nucleus. *Trends Cell Biol.* **2011**, *21*, 336–343. [[CrossRef](#)] [[PubMed](#)]
165. Pierce, B.L.; Tong, L.; Argos, M.; Demanelis, K.; Jasmine, F.; Rakibuz-Zaman, M.; Sarwar, G.; Islam, M.T.; Shahriar, H.; Islam, T.; et al. Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. *Nat. Commun.* **2018**, *9*, 804. [[CrossRef](#)] [[PubMed](#)]
166. Arnold, C.D.; Gerlach, D.; Spies, D.; Matts, J.A.; Sytnikova, Y.A.; Pagani, M.; Lau, N.C.; Stark, A. Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* **2014**, *46*, 685–692. [[CrossRef](#)] [[PubMed](#)]
167. Weber, M.; Hellmann, I.; Stadler, M.B.; Ramos, L.; Pääbo, S.; Rebhan, M.; Schübeler, D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **2007**, *39*, 457–466. [[CrossRef](#)] [[PubMed](#)]
168. Luscombe, N.M.; Babu, M.M.; Yu, H.; Snyder, M.; Teichmann, S.A.; Gerstein, M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **2004**, *431*, 308–312. [[CrossRef](#)]
169. Wittkopp, P.J. Genomic sources of regulatory variation in cis and in trans. *Cell. Mol. Life Sci.* **2005**, *62*, 1779–1783. [[CrossRef](#)]
170. Yang, B.; Wittkopp, P.J. Structure of the Transcriptional Regulatory Network Correlates with Regulatory Divergence in *Drosophila*. *Mol. Biol. Evol.* **2017**, *34*, 1352–1362. [[CrossRef](#)]
171. Kittelmann, S.; Buffry, A.D.; Franke, F.A.; Almudi, I.; Yoth, M.; Sabaris, G.; Couso, J.P.; Nunes, M.D.S.; Frankel, N.; Gómez-Skarmeta, J.L.; et al. Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution. *PLoS Genet.* **2018**, *14*, e1007375. [[CrossRef](#)] [[PubMed](#)]
172. McGregor, A.P.; Orgogozo, V.; Delon, I.; Zanet, J.; Srinivasan, D.G.; Payre, F.; Stern, D.L. Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* **2007**, *448*, 587–590. [[CrossRef](#)] [[PubMed](#)]
173. Arif, S.; Murat, S.; Almudi, I.; Nunes, M.D.S.; Bortolamiol-Becet, D.; McGregor, N.S.; Currie, J.M.S.; Hughes, H.; Ronshaugen, M.; Sucena, É.; et al. Evolution of mir-92a underlies natural morphological variation in *Drosophila melanogaster*. *Curr. Biol.* **2013**, *23*, 523–528. [[CrossRef](#)]
174. Khaitovich, P.; Hellmann, I.; Enard, W.; Nowick, K.; Leinweber, M.; Franz, H.; Weiss, G.; Lachmann, M.; Pääbo, S. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **2005**, *309*, 1850–1854. [[CrossRef](#)] [[PubMed](#)]
175. Yang, J.; Su, A.I.; Li, W.-H. Gene expression evolves faster in narrowly than in broadly expressed mammalian genes. *Mol. Biol. Evol.* **2005**, *22*, 2113–2118. [[CrossRef](#)] [[PubMed](#)]
176. Amaral, A.J.; Brito, F.F.; Chobanyan, T.; Yoshikawa, S.; Yokokura, T.; van Vactor, D.; Gama-Carvalho, M. Quality assessment and control of tissue specific RNA-seq libraries of *Drosophila* transgenic RNAi models. *Front. Genet.* **2014**, *5*, 43. [[CrossRef](#)]
177. Potier, D.; Davie, K.; Hulselmans, G.; Naval Sanchez, M.; Haagen, L.; Huynh-Thu, V.A.; Koldere, D.; Celik, A.; Geurts, P.; Christiaens, V.; et al. Mapping gene regulatory networks in *Drosophila* eye development by large-scale transcriptome perturbations and motif inference. *Cell Rep.* **2014**, *9*, 2290–2303. [[CrossRef](#)]
178. Hickman, S.E.; Kingery, N.D.; Ohsumi, T.K.; Borowsky, M.L.; Wang, L.-c.; Means, T.K.; El Khoury, J. The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* **2013**, *16*, 1896–1905. [[CrossRef](#)]
179. Ahn, R.S.; Taravati, K.; Lai, K.; Lee, K.M.; Nititham, J.; Gupta, R.; Chang, D.S.; Arron, S.T.; Rosenblum, M.; Liao, W. Transcriptional landscape of epithelial and immune cell populations revealed through FACS-seq of healthy human skin. *Sci. Rep.* **2017**, *7*, 1343. [[CrossRef](#)]
180. Florio, M.; Albert, M.; Taverna, E.; Namba, T.; Brandl, H.; Lewitus, E.; Haffner, C.; Sykes, A.; Wong, F.K.; Peters, J.; et al. Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **2015**. [[CrossRef](#)]

181. Karaiskos, N.; Wahle, P.; Alles, J.; Boltengagen, A.; Ayoub, S.; Kipar, C.; Kocks, C.; Rajewsky, N.; Zinzen, R.P. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **2017**, *323*, eaan3235. [[CrossRef](#)]
182. Seb e-Pedr s, A.; Saudemont, B.; Chomsky, E.; Plessier, F.; Mailh e, M.-P.; Renno, J.; Loe-Mie, Y.; Lifshitz, A.; Mukamel, Z.; Schmutz, S.; et al. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* **2018**, *173*, 1520–1534. [[CrossRef](#)] [[PubMed](#)]
183. Fincher, C.T.; Wurtzel, O.; de Hoog, T.; Kravarik, K.M.; Reddien, P.W. Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **2018**, *360*. [[CrossRef](#)] [[PubMed](#)]
184. Achim, K.; Eling, N.; Vergara, H.M.; Bertucci, P.Y.; Musser, J.; Vopalensky, P.; Brunet, T.; Collier, P.; Benes, V.; Marioni, J.C.; et al. Whole-Body Single-Cell Sequencing Reveals Transcriptional Domains in the Annelid Larval Body. *Mol. Biol. Evol.* **2018**, *35*, 1047–1062. [[CrossRef](#)] [[PubMed](#)]
185. Davie, K.; Janssens, J.; Koldere, D.; de Waegeneer, M.; Pech, U.; Kreft, L.; Aibar, S.; Makhzami, S.; Christiaens, V.; Bravo Gonz lez-Blas, C.; et al. A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain. *Cell* **2018**. [[CrossRef](#)] [[PubMed](#)]
186. Konstantinides, N.; Kapuralin, K.; Fadil, C.; Barboza, L.; Satija, R.; Desplan, C. Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. *Cell* **2018**. [[CrossRef](#)]
187. Croset, V.; Treiber, C.D.; Waddell, S. Cellular diversity in the *Drosophila* midbrain revealed by single-cell transcriptomics. *eLife* **2018**, *7*. [[CrossRef](#)]
188. Aibar, S.; Gonz lez-Blas, C.B.; Moerman, T.; Huynh-Thu, V.A.; Imrichova, H.; Hulselmans, G.; Rambow, F.; Marine, J.-C.; Geurts, P.; Aerts, J.; et al. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **2017**, *14*, 1083–1086. [[CrossRef](#)]
189. Chan, T.E.; Stumpf, M.P.H.; Babbie, A.C. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst.* **2017**, *5*, 251–267. [[CrossRef](#)]
190. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)]
191. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [[CrossRef](#)] [[PubMed](#)]
192. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)] [[PubMed](#)]
193. Mootha, V.K.; Lindgren, C.M.; Eriksson, K.-F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstr le, M.; Laurila, E.; et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **2003**, *34*, 267–273. [[CrossRef](#)] [[PubMed](#)]
194. Goeman, J.J.; B hlmann, P. Analyzing gene expression data in terms of gene sets: Methodological issues. *Bioinformatics* **2007**, *23*, 980–987. [[CrossRef](#)] [[PubMed](#)]
195. Nam, D.; Kim, S.-Y. Gene-set approach for expression pattern analysis. *Brief. Bioinform.* **2008**, *9*, 189–197. [[CrossRef](#)] [[PubMed](#)]
196. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.-Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316–22. [[CrossRef](#)] [[PubMed](#)]
197. Wang, S.; Cao, M.; Ma, X.; Chen, W.; Zhao, J.; Sun, C.; Tan, L.; Liu, F. Integrated RNA Sequencing and QTL Mapping to Identify Candidate Genes from *Oryza rufipogon* Associated with Salt Tolerance at the Seedling Stage. *Front. Plant Sci.* **2017**, *8*, 1427. [[CrossRef](#)] [[PubMed](#)]
198. Jian, H.; Zhang, A.; Ma, J.; Wang, T.; Yang, B.; Shuang, L.S.; Liu, M.; Li, J.; Xu, X.; Paterson, A.H.; et al. Joint QTL mapping and transcriptome sequencing analysis reveal candidate flowering time genes in *Brassica napus* L. *BMC Genom.* **2019**, *20*, 21. [[CrossRef](#)] [[PubMed](#)]
199. Zhang, Z.; Jia, Y.; Almeida, P.; Mank, J.E.; van Tuinen, M.; Wang, Q.; Jiang, Z.; Chen, Y.; Zhan, K.; Hou, S.; et al. Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *GigaScience* **2018**, *7*. [[CrossRef](#)]

