Peer-reviewed version available at *Genes* **2019**, *10*, 492; doi:10.3390/genes10070492

Review

Cloudy with a chance of insights: Context dependent gene regulation and implications for the evolution of gene expression

Elisa Buchberger[1#], Micael Reis[1#], Ting-Hsuan Lu[1,2], Nico Posnien[1*]

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

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

[#]equal contribution

[*]contact: nposnie@gwdg.de

Abstract

Research in various fields of evolutionary biology has shown that divergence in gene expression is a key driver for phenotypic variation. An exceptional contribution of *cis*-regulatory evolution has for instance 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 developmental stage-specific expression. We argue that the regulatory context must be considered when studying evolution of gene expression.

Keywords

gene expression, gene regulation, evolution, allele specific expression, eQTL, RNAseq, ChIPseq, chromatin, ATACseq, genotype-phenotype map

Peer-reviewed version available at Genes 2019, 10, 492; doi:10.3390/genes10070492

Introduction

Living organisms are uniquely characterized by their appearance, their function as well as 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 a fundamental process of 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 differences in DNA sequences and phenotypic variation 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 difference and phenotypic variation. Since these non-coding regions may contain important regulatory sequences, which interact with DNA-binding proteins as well as non-coding RNAs involved in gene regulation, 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].

In this review we will first summarize findings illustrating the importance of gene expression divergence in phenotypic evolution for various traits such as morphology, behavior, physiology and life-history. Next, we will review various mechanisms underlying gene regulation and we will highlight how they facilitate context dependent gene expression. Current approaches aiming at understanding genome-wide patterns of gene expression divergence as well as the underlying molecular mechanisms will be presented. Eventually, we will summarize limitations of current approaches in the light of context dependent gene regulation and we will suggest how various datasets can be integrated to gain comprehensive insights into the evolution of gene expression.

Gene expression divergence and phenotypic evolution

Changes in gene expression have been linked to many phenotypes studied so far. 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 molecular variation affecting gene expression divergence with phenotypes are based on studying simple morphological traits, such as the evolution of trichome patterns in *Drosophila* or variation in body coloration. For instance, a clear link between changes in the regulatory region of the *shavenbaby* gene and variation in trichome patterns across *Drosophila* species has been established [22,23]. Similarly, individual nucleotide polymorphisms in the *ebony* [24] and *yellow* genes [25] underly natural variation 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 variation in the expression of *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. Variation 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 both genitalia bristle and sex comb sensory teeth number [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 expression differences of the gene *syntaxin1a*, since higher expression of this gene is directly correlated with a social life style in bees [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*) has been shown to be influenced by differential expression of the gene *vasopressin* [32]. These examples impressively demonstrate that natural variation in different behavioral traits is associated with divergence in gene expression.

Many studies exploring the molecular basis of differences in 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 differences of a single gene (*PtFT2*) are responsible for 65% of the variation in timing of

bud set [35]. Other studies similarly identified mutations in *cis*-regulatory regions causing variation in gene expression which ultimately affects an organism's physiological response to the environment. For example, a 2 bp deletion in the promoter region of gene *ERG28* in *Saccharomyces cerevisiae* has been shown to result 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, variation in gene expression is a major driver for phenotypic divergence of morphological and behavioural traits as well as for life history and physiological traits.

Gene expression and gene regulation are highly context dependent

The proper function of an organism relies on the correct expression of genes and the interplay of gene products in each of its organs. Depending on the life span of the organism, each cell in an adult organ must fulfil distinct functions for a long period of time. These functions are ensured by tissue or even cell-specific gene expression. It has, for instance, been shown that the fundamental function of the brain to form a long-term memory based on environmental cues and experiences is associated with cell-type specific changes in gene expression profiles in *Drosophila melanogaster* [38]. Disturbance of these fundamental functions ultimately leads to diseases and eventually to the death of the respective organism. In humans, cancer formation and progression are strongly linked to changes in the transcriptional profiles in the affected tissue [39]. Therefore, the expression of genes must be tightly regulated to ensure allocation of the correct gene products at the right time in the right cells.

While it is broadly accepted that gene expression and thus gene regulation is highly context dependent [40], the underlying molecular mechanisms are being revealed only recently. In the following we will highlight some major molecular mechanisms facilitating context dependent gene regulation. Since a lot of our current understanding about the regulation of gene expression on an organismic scale has been deduced from developmental genetics studies, the following examples will be biased towards this research field. Of course, this does not interfere with the generality of the presented findings.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2019 doi:10.20944/preprints201905.0330.v1

Peer-reviewed version available at Genes 2019, 10, 492; doi:10.3390/genes10070492

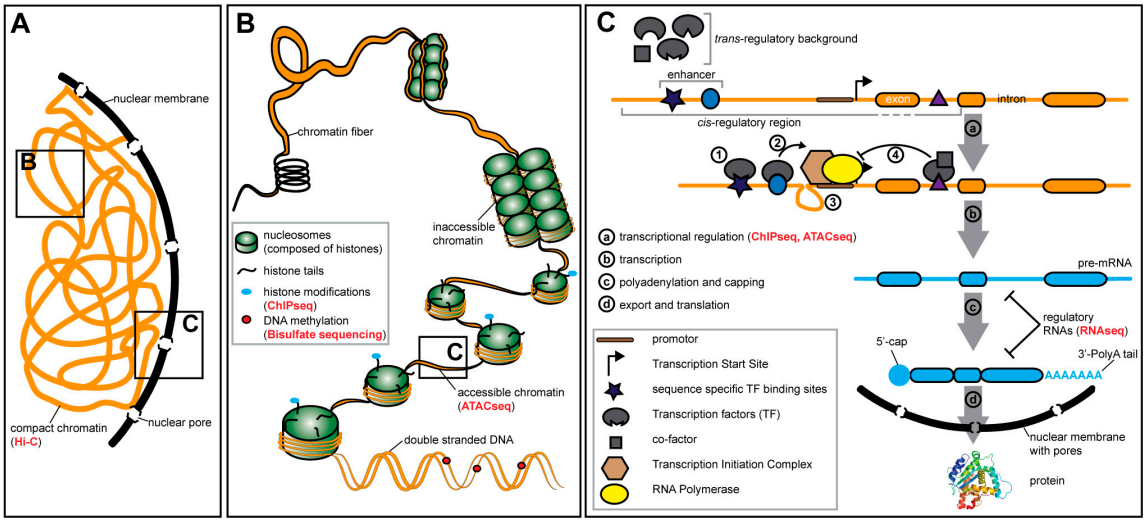

Figure 1. Gene expression is regulated on various levels. (**A**) The DNA is compressed in the nucleus of the cell. (**B**) The DNA in the nucleus is compressed by binding of histone proteins. The chromatin contains easily accessible euchromatin regions and highly compact and inaccessible heterochromatin 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 in (looping) of the DNA (3). TFs can also repress transcription, for instance by binding of a co-factor (4). 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.

129

130 *Pre-transcriptional regulation - chromatin states and methylation*

131 First regulatory mechanisms are at play on the level of genome organisation. Compressed
132 DNA in the nucleus forms a tertiary structure (Figure 1A) that can be studied in detail by an NGS
133 based chromosome conformation capture method called Hi-C [41] (Table 1). Hi-C applied in various
134 bilaterians revealed one fundamental characteristic of the genome: Some regions of the genome
135 interact consistently more often than other regions [42–44]. These topologically associating domains
136 (TADs) have been shown to influence gene expression. For instance, the famous temporal and spatial
137 collinearity of *Hox* gene expression in the developing vertebrate limb has been associated with the
138 location of the *HoxD* cluster in a gene desert that lies between two adjacent TADs [45]. Chromatin
139 interactions between promoter regions and distant regulatory sequences are pervasive in vertebrates.
140 The application of Hi-C in different human primary blood cell types showed that these interactions
141 are highly cell-type specific [46]. How the three-dimensional organization of the genome exactly
142 influences gene regulation has just started to be revealed and represents an active and exciting field
143 of research.

144 Much more detailed knowledge has been accumulated on the level of chromatin organization
145 and its impact on gene expression. The compressed DNA in the nucleus can be subdivided into
146 loosely packed and easily accessible euchromatin and condensed and inaccessible heterochromatin
147 (Figure 1B). Most actively transcribed genes are located in euchromatic regions and their DNA
148 sequence is free of nucleosomes to allow transcription factors to bind. 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) [47] (Table 1). Recent application of ATACseq on single cells originating from 13 different mouse tissues [48] and from three stages of *Drosophila* embryonic development [49] revealed clear signatures of cell type and stage specific chromatin accessibility states. The chromatin state and thus DNA accessibility is influenced by modifications of histone proteins, the subunits of nucleosomes. A clear link between histone modifications and gene regulation has been established [50,51] and genome wide surveys based on ChIPseq (Chromatin Immunoprecipitation followed by next generation DNA sequencing) [52,53] (Table 1) revealed tissue and cell type specific histone modification patterns [54,55].

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 [56] and they can recruit proteins of the Polycomb Repressive Complex 2 (PRC2) and histone methyltransferases, which are associated with inactive heterochromatin. Intriguingly, the lncRNA-protein interactions are tissue specific since they were observed in mouse placenta cells, but not in liver tissue [57]. lncRNAs have also been implicated in transcriptional activation via direct interaction with a protein complex that mediates histone H3 lysine 4 trimethylation, an active histone mark [58]. Transcriptional activation is also achieved through direct interaction of lncRNAs with protein complexes that stabilize enhancer-promotor interactions [59]. 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 [60]. Since miRNAs are highly tissue specifically expressed [61–63], these molecules provide an excellent mechanism to facilitate tissue specific chromatin accessibility.

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 [64,65] and recent data has shown that transcription factors can integrate methylation patterns to refine gene regulation [66]. Since the methylation is highly dynamic, for instance throughout cellular differentiation [67], it facilitates context dependent gene regulation.

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 modulate gene expression pre-transcriptionally.

Transcriptional regulation – transcription factors and cis-regulatory elements

Once the chromatin is accessible for proteins, gene expression is directly regulated by protein-DNA interactions [68,69] (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 [70]. 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 [71] and the regional identity of neuroblasts is regulated by the expression of spatially restricted transcription factors [72]. 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 [73,74]. 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 [75,76]. Please note that *cis*-regulatory elements do not only have positive effects on gene expression. *cis*-regulatory elements with a repressive impact on gene expression (i.e. silencers) were described, as well as elements protecting a given gene locus from adjacent regulatory input (i.e. insulators). Although we focus here on enhancers, many of the discussed aspects apply to these elements as well. Enhancers have been shown to be highly modular [77,78]. This is best 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 defined by five enhancers with each of them being responsible for an individual stripe or a pair of stripes [79,80]. During *Drosophila* eye development, the gene *eyes-absent* (*eya*) is expressed in undifferentiated cells that undergo cell division, as well as in differentiating cells that do not divide anymore. Although these cells are present in the same tissue in close proximity, the expression of *eya* is regulated by different combinations of several enhancers [81]. The modularity and dynamic usage of enhancers have also been shown on a genome-wide scale during *Drosophila* embryonic development [82]. 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 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 of the wing imaginal disc, while the presence and binding of its co-factor U-shaped (Ush) transforms it into a transcriptional repressor in adjacent regions [83–85]. 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 [86]. 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 defined DNA sequences specifies the unique transcriptional landscape of a developing cell or cell groups.

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 [87–89]. 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 [90] 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 [91].

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) [92] or they are part of RNA-protein complexes (e.g. lncRNA) where they influence gene regulation on various levels [93]. lncRNAs present in the cytoplasm also influence mRNA stability [94,95] and they can protect mRNA against targeted degradation by trapping miRNAs in a sponge-like mechanisms [96]. Regulatory RNAs play a major role during development [93,97] and their expression has been shown to be cell type specific [98,99]. 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).

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

Method	Key information
RNAseq	**Summary:** RNA is isolated and reversed transcribed into cDNA for library preparation and sequencing. **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 [100]. 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) [101]. For small RNA enrichment several commercial kits are available which select for molecule sizes less than 30 nucleotides [102]. **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) [103,19,18,20,21]. RNAseq is also a useful tool for miRNA profiling and annotation [104] as well as differential expression of lncRNAs [105]. **Single cell application:** [106–108]
ATACseq*	**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 [47,109]. This method substituted previous ones such as DNaseseq and FAIREseq, due to its simplicity and effectiveness. **Practical considerations:** Usually the protocol should be done with fresh tissue and a defined number of nuclei/cells (e.g 500-50000 for mammalian tissues [109]) 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 [110]. **Applications:** ATACseq is commonly used to complement RNAseq data to identify potential regulatory regions (enhancers) [111]. 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 [47]. **Single cell application:** [112,113]
ChIPseq*	**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. Upon reversal of the crosslink and dissociation of the DNA libraries for short read sequencing are prepared [52,53]. **Practical considerations:** This technique relies on previous knowledge about the DNA-binding proteins and available antibodies. **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 [114]. It has been used also to estimate histone modifications and nucleosome position between different species [54]. **Single cell application:** [115]
Hi-C*	**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 [116]. **Practical considerations:** Hi-C relies in restriction enzyme recognition sites which can create bias due to their heterogeneous distribution in the genome [117]. Alternative methods used DNase I [118] or micrococcus digestion [119] to overcome that issue. **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 [120,42,121]. **Single cell application:** [122]
BSseq	**Summary**: DNA is treated with sodium bisulfite to deaminate cytosine bases into uracil (thymine after PCR) while methyl-cytosine bases are not affected [123]. The treated DNA is then digested for library preparation and sequencing [124]. **Practical considerations:** The deamination reaction usually has high yield, but small variations can create significant bias in the estimation of global methylation patterns [125]. 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 126]. **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 [124]. In recent years, this method has been extensively applied to ecological and evolutionary studies [127,125]. **Single cell application:** [128,129]

255

256

Peer-reviewed version available at Genes 2019, 10, 492; doi:10.3390/genes10070492

Evolution of gene expression is context dependent

Genome wide expression studies have been broadly used to show that the evolution of gene expression is context dependent. A study of six homologous organs in nine mammals and one bird showed that gene expression evolves at different speeds in different tissues as well as in different lineages. While gene expression was more stable in nervous system tissue, it evolved more rapidly in testes. Similarly, this work revealed that gene expression variation was less pronounced in rodents compared to apes [130]. Additionally, it has been shown that genes that are expressed in many tissues, i.e. pleiotropic factors, experience more constraints than tissue specific genes [131,132]. Comparative expression studies have also been employed to assess the impact of developmental stages on the evolution of gene regulation. The analysis of expression data from various developmental stages in different vertebrates revealed the pharyngula stage to be most constraint (i.e. most similar) [133,134]. 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 [135].

Comparative studies in five mammals [136] and twelve *Drosophila* species [62] revealed that also non-coding RNA molecules evolve in a highly context dependent manner. For instance, high turnover rates have been associated with the underlying molecular mechanisms involved in miRNA maturation [62]. Additionally, novel miRNAs tend to be expressed highly tissue specifically [136,62], suggesting that the evolvability of miRNAs is constraint by the regulatory context. Structural constraints on sequence variability have also been observed for lncRNAs, since they are mostly conserved across vertebrates in 5-prime exonic regions [137].

These examples clearly demonstrate that the context in which gene expression variation is studied, such as the type of tissue or developmental stage poses constraints on the overall evolvability of gene expression.

Evolutionary correlation studies in the light of context dependent gene expression

Many studies employ gene expression as intermediate phenotype to link genetic variation to trait divergence. Since high throughput sequencing methods are highly sensitive, even subtle changes in gene expression can be detected. Therefore, these analyses commonly result in large lists of genes that show significant differential expression between phenotype classes. In the following we highlight some characteristics of such gene lists and we will present ideas to improve the identification of meaningful candidate genes in the light of context dependent gene expression.

Impact of tissue-specificity on signal-to-noise ratio

Not all identified genes may be directly associated with the trait of interest, but rather represent background noise. If the analyzed tissue or time point has not been chosen as specifically as possible (Table 2), it is more difficult to separate background signal from putative relevant signals. Few studies specifically tested whether complex tissue composition indeed influences the sensitivity to detect gene expression differences. A RNAseq study in *Drosophila melanogaster* compared genome wide expression of wildtype central nervous system tissue to tissue extracted from 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 [138]. 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 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 [139–141] as well as to reveal candidate genes in evolutionary studies [142]. 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 [106–108] (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* [143], the cnidarian *Nematostella vectensis* [144], the planarian *Schmidtea mediterranea* [145] or the marine annelid *Platynereis dumerilii* [146]. 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 [147], but also for parts of the brain such as the optic lobes [148] and the mid-brain [149]. Eventually, scRNAseq also allows to gain new major insights into developmental processes, such as cell lineage specification [150], regulation of cell differentiation [151], molecular underpinnings of pluripotency [152] and the reconstruction of cell specific gene regulatory networks [153,154]. In summary, scRNAseq represents an exciting new method to study context dependent gene expression and regulation in complex tissues.

Table 2. Comparison of different RNA sequencing methods.

	bulkRNAseq of whole individuals	bulkRNAseq 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	-	+/-	+

Integration of information on biological functions helps identifying meaningful candidate 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 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 [155,156] 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) [157–160] can be employed to reveal if specific molecular or developmental pathways are involved in the development of the trait of interest [161]. Hence, the implementation of biological knowledge helps finding patterns in an otherwise unstructured dataset and helps to restrict the number of meaningful candidates.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 May 2019 doi:10.20944/preprints201905.0330.v1

Peer-reviewed version available at Genes 2019, 10, 492; doi:10.3390/genes10070492

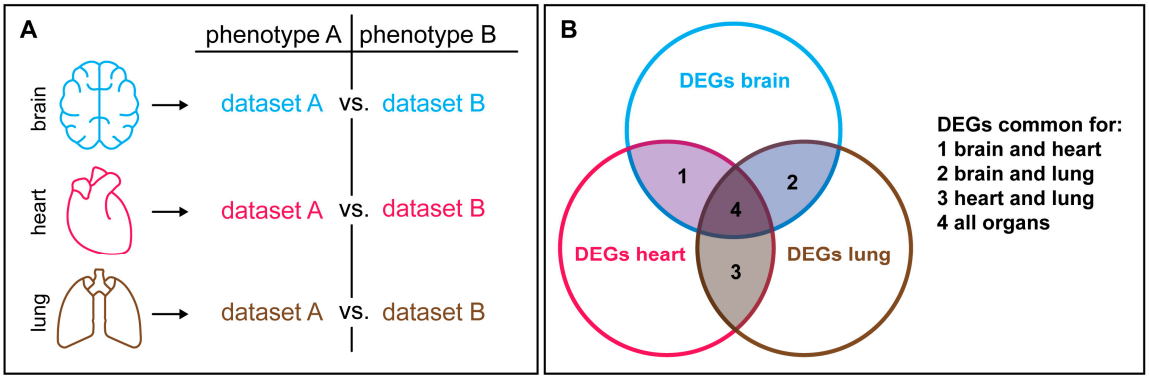

Figure 2. Generic factors that across different tissue can be excluded in correlation studies. (**A**) If specific candidate genes that are differentially expressed between phenotype A and B are supposed to be revealed, one can generate a comparable dataset for additional tissues. (**B**) Each pairwise 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.

348 *Towards an integrative approach to establish genotype-phenotype associations*

349 Another important source of reduction of candidate genes in correlative studies is the
350 generation of complementary datasets. Quantitative genetics approaches, such as genome wide
351 association studies (GWAS) or quantitative trait loci (QTL) mapping can be applied to identify
352 candidate loci that show association between phenotypic divergence and genomic variation. This
353 data can be extremely helpful to overlap with candidate gene lists obtained from differential gene
354 expression studies. This combinatorial approach has been successfully applied to reveal candidate
355 genes involved in different nest building behaviors among the two mouse species *Peromyscus*
356 *polionotus* and *P. maniculatus*. The 498 candidate genes in a QTL region associated with behavioral
357 differences were reduced to 23 genes that were differentially expressed in the brain region
358 responsible for the studied behavior. Only nine of these 23 genes showed signatures of *cis*-regulatory
359 divergence, suggesting that these were excellent candidate genes. Subsequent functional validation
360 tests confirmed the involvement of one of the candidates in the respective behavioral differences
361 [32]. Similar combinations of QTL mapping and genome wide differential expression analyses have
362 been applied to identify key candidate genes responsible for variation in salt tolerance in rice (*Oryza*
363 *rufipogon*) [162] and flowering time in rape (*Brassica napus*) [163].

364 Also, population genetics data that provides genome wide insights into signatures of
365 selection may be meaningful to overlap with expression data. This has been successfully used in
366 domestication studies. The combination of two datasets revealed one gene that was differentially
367 expressed between wild and domestic duck brains and showed signs of selection. This gene is a
368 strong candidate that may explain phenotypic differences associated with duck domestication [164].
369 These examples show that the combination of independent datasets providing information about
370 different levels of the genotype-phenotype map are thus powerful ways to link genetic variation to
371 phenotypic divergence.

372

Peer-reviewed version available at Genes 2019, 10, 492; doi:10.3390/genes10070492

Evolution of gene regulation mechanisms

The accumulation of comparative gene expression data triggered a strong interest in unravelling the molecular and evolutionary mechanisms underlying divergence in gene expression. 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 QTL (eQTL) mapping and allele-specific expression studies (ASE). In the following, we summarize the key concepts of the two approaches, and we discuss limitations and chances in the light of context dependent gene expression and regulation.

eQTL and ASE studies to reveal mechanisms underlying evolution of gene expression and gene regulation

eQTL studies are basically QTL or 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 [165–167]. Therefore, normal QTL or association mapping methodology can be applied to reveal genomic variants for genome wide expression divergence. A typical outcome of such studies is the identification of genomic variants that are either located close to the gene that shows expression divergence (local QTLs; also referred to as *cis*-QTLs) or far away (distant QTLs; also referred to as *trans*-QTLs). It is important to note here that the distinction is purely based on the location of the variant with respect to the studied gene and that *cis* and *trans* in this context does not refer to any mechanism involved (see below). eQTL studies already successfully revealed some fundamental principles underlying gene expression divergence. Various studies have shown that gene expression is indeed highly variable across individuals and heritability estimates support the contribution of a genetic component [168].

While eQTL studies reveal genomic loci or individual SNPs associated with expression difference, ASE studies in F1 hybrids represent a powerful approach to gain mechanistic insights into differential gene expression [169]. 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) [170,171]. *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. ASE studies in various organisms contributed to exciting general insights into mechanisms underlying gene expression divergence. The most

consistent observation is that *cis*-regulatory divergence seems to be prevalent in intra- as well as interspecific comparisons [172–175]. Exceptions have been observed for instance for comparisons between the cosmopolitan fly species *D. melanogaster* and the closely related specialist species *D. sechellia* [176]. Nevertheless, 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.

Limitations of eQTL and ASE studies

Although eQTL and ASE studies revealed fundamental concepts underlying gene expression variation, a comprehensive molecular understanding is still lacking. An intrinsic limitation of eQTL studies for instance is to ascribe a specific regulatory mechanism (see Figure 1) to the identified genetic variants. While it is conceivable that variation close to the gene locus (i.e. *cis*-QTLs) may affect expression of the gene due to differences in promotor by modification of transcription factor-DNA binding, it is much more complicated to assign regulatory mechanisms to *trans*-QTLs. Similarly, ASE studies alone do not allow identifying specific genetic variants. 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 [177], *Arabidopsis thaliana* [178], maize [179] and *Drosophila* [176,180]. A combination of ASE and SNP data obtained from lymphoblastoid cell lines from the 1,000 Genomes Project further strongly suggests that genetic variation is a common explanation for allele-specific gene expression [181]. As for eQTL studies, it is much more complicated to define *trans*-regulatory divergence mechanistically since these effects can result from various factors, such as expression or coding differences of transcription factors, the presence of transcriptional co-factors or the differential expression of regulatory RNA molecules.

In addition to the complication to assign regulatory mechanisms, eQTL studies have shown that the level of heritability is relatively low and the effect size of identified associated genomic variants is normally small [182]. These observations strongly suggest that gene expression divergence is affected by various other factors, such as epigenetic modifications or environmental cues. Therefore, a much better integration of complementary datasets is needed. We argue that the combination of eQTL and ASE studies with comparative datasets aiming at identifying open chromatin regions, epigenetic features or regulatory RNAs has the potential to identify new molecular mechanisms driving the evolution of gene expression.

From eQTLs and ASE to mechanisms – natural variation in regulatory traits

Understanding the evolution of gene expression on a mechanistic level requires revealing how natural variation at the different levels of gene regulation (see Figure 1) influences gene expression. Exceptional insights into the impact of genetic variation associated with different regulatory traits, such as chromatin architecture, chromatin accessibility, histone modifications as well as alternative splicing on gene expression and the resulting protein composition have been gained by studying lymphoblastoid cell lines as part of the HapMap2 [183] and 1000 Genomes Project [181]. These 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 and eQTLs.

eQTL studies revealed SNPs affecting all levels of gene regulation including genome organization [184], chromatin accessibility [185], histone modifications, RNA-Polymerase II occupancy and eventually gene expression [186,187]. An exciting link has been established between variation in transcription factor binding affinity and differences in histone modifications. Genetic variants that confer higher transcription factor binding affinity are also associated with an increase in active histone marks [186]. This observation has been functionally validated for the RE1-silencing transcription factor (REST) during neural differentiation in mouse. It has been shown that REST recruits the repressive histone mark H3K27me3 which is depleted upon loss of REST function or loss of the REST binding motifs. Intriguingly, the artificial insertion of REST binding motifs is sufficient to recruit ectopic H3K27me3 [188], supporting a tight causal link between transcription factor binding and histone modifications. 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 [189], it is conceivable that natural genetic variation may very quickly affect gene expression on various levels ranging from transcription factor binding to histone modification and chromatin accessibility.

Since about 65% of the eQTLs (i.e. variation in gene expression) are associated with histone modifications and chromatin accessibility [190], the final composition of the proteome must be additionally controlled by other processes, such as post-transcriptional regulation. Most importantly, mRNA splicing contributes to the complexity of the protein content in a cell and individual SNPs have been associated with differences in splicing [191,190]. These so-called splicing QTLs (sQTLs) were predominantly found at the respective splice sites close to exon boundaries (i.e. *cis* sQTLs) and some of the identified *trans* sQTLs were linked to proteins involved in RNA processing [191,190]. Since the spliceosome is already assembled during ongoing transcription, the chromatin state and the transcription rate can influence splicing events [192]. The observation that CTCF binding, which is highly influenced by DNA methylation patterns, results in RNA-Polymerase II pausing and subsequent inclusion of weak upstream exons have started to reveal the mechanistic interplay of epigenetic marks and differential splicing [193].

GWAS studies of whole blood samples from 347 human individuals have also revealed a link between natural variation of epigenetic methylation patterns and gene expression, by suggesting 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 [194]. Interestingly, 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 [195]. In summary, variation in gene expression, gene regulation and methylation are tightly linked and differences in methylation levels can shape the genetic composition.

The outlined findings provide exceptional mechanistic insights into the effect of genetic variants on all levels of gene regulation and thus gene expression. Especially, the observation that many regulatory traits are functionally linked and therefore similarly affected by SNPs may explain why natural variation in gene expression is pervasive in nature and can drive phenotypic diversification.

Context dependent gene expression and the evolution of gene regulation

Although the association data obtained from human cell lines allow unprecedented detailed insights into the impact of natural variation on gene regulation, they only represent a snapshot of the complexity of gene expression. This is largely due to the context dependence of gene expression and gene regulation. While phenotypic traits in classical association or linkage mapping studies are clearly defined, gene expression varies significantly across developmental stages, tissues or even cell types. 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 [196]. Similarly, the analysis of sexually dimorphic gene expression in different organs in intercrosses of two inbred mouse strains revealed that the expression of dimorphic genes was tissue specific. Additionally, tissue specific eQTL regions were identified, suggesting that expression differences between sexes are regulated by tissue specific regulatory elements [197].

ASE studies have identified general mechanisms underlying the evolution of gene expression. However, to date, ASE studies mainly provide insights into expression divergence in adults and for entire organisms. Indeed, very few studies have used this approach to evaluate divergence types in different conditions. It has for instance been shown in closely related *Arabidopsis* species that environmental factors influence the excess of *cis*-regulatory divergence [198,199]. Similarly, different contributions of *cis*- and *trans*-divergence (or combinations thereof) have been obtained by comparing differently aged flies [200] and when data from entire fly bodies was compared to heads only [174,176]. A recent ASE study using tissue specific data for Malpighian

514 tubules of different *D. melanogaster* populations further supports the need for more defined analyses
515 [201]. In the light of context dependent gene regulation these first results call for an integration of
516 stage or tissue specific aspects of gene expression in eQTL and ASE studies in order to reveal
517 whether patterns observed so far will hold true across highly variable regulatory environments.

518

519 **Outlook and Summary**

520 We are currently living in exciting times with functional genomics and transcriptomics
521 methods being widely applicable (Table 1). This provides an unprecedented opportunity to test,
522 whether mechanistic insights obtained by highly coordinated consortia studying human cell lines as
523 well as tractable genetic model system such as yeast and *Drosophila* hold true in other study systems.
524 For instance, epigenetic factors are being studied with respect to transgenerational information
525 transfer [202] and in relation to phenotypic plasticity [203] in various emerging model systems.
526 Therefore, genome wide information on methylation patterns [204] should be studied in many
527 systems in the context of gene regulation. Chromatin accessibility can be studied applying ATACseq
528 [47] in many systems to relate comparative gene expression data or results obtained from ASE studies
529 to genomic regions that most likely contain meaningful regulatory elements. Similarly, ChIPseq
530 [52,53] for certain histone modifications that either mark active or inactive gene loci [205] is
531 applicable in a variety of organisms. Another exciting new area of research deals with the impact of
532 the three-dimensional organization of the genome on gene expression and regulation. Few studies
533 compared genome organization across organisms [206], but the picture emerges that global
534 organizational patterns (e.g. TADs) may be highly conserved, suggesting that the three-dimensional
535 chromatin organization can pose constraints on the evolution of gene regulation.

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

546

Peer-reviewed version available at Genes 2019, 10, 492; doi:10.3390/genes10070492

Author Contributions

Conceptualization, all authors, writing—review and editing, all authors; visualization, NP; supervision, NP

Funding

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

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, doi:10.1038/nrg1726.

2. Hatefi, Y. The mitochondrial electron transport and oxidative phosphorylation system. *Annu. Rev. Biochem.* **1985**, *54*, 1015–1069, doi:10.1146/annurev.bi.54.070185.005055.

3. Saraste, M. Oxidative Phosphorylation at the fin de siècle. *Science* **1999**, *283*, 1488–1493, doi:10.1126/SCIENCE.283.5407.1488.

4. Janeway, C.A.; Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **2002**, *20*, 197–216, doi:10.1146/annurev.immunol.20.083001.084359.

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, doi:10.1038/nri.2016.29.

6. Lumpkin, E.A.; Caterina, M.J. Mechanisms of sensory transduction in the skin. *Nature* **2007**, *445*, 858–865, doi:10.1038/nature05662.

7. Yau, K.-W.; Hardie, R.C. Phototransduction motifs and variations. *Cell* **2009**, *139*, 246–264, doi:10.1016/j.cell.2009.09.029.

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. *Journal of Experimental Medicine* **1944**, *79*, 137–158, doi:10.1084/jem.79.2.137.

9. THODAY, J.M. Location of Polygenes. *Nature Insight Biodiverstiy* **1961**, *191*, 368–370, doi:10.1038/191368a0.

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. *Evolution; International Journal of Organic Evolution* **2007**, *61*, 1017–1032, doi:10.1111/j.1558-5646.2007.00100.x.

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, doi:10.1038/nature10811.

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, doi:10.1126/science.1233213.

14. Hoekstra, H.E.; Hirschmann, R.J.; Bundey, 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, doi:10.1126/science.1126121.

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, doi:10.1126/science.1224829.

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

598 sequence quantitatively alter sensory organ development and function. *eLife* **2017**, *6,*
599 *doi:*10.7554/eLife.26402.

600 17. Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics.
601 *Nature reviews. Genetics* **2009**, *10*, 57–63, doi:10.1038/nrg2484.

602 18. Oppenheim, S.J.; Baker, R.H.; Simon, S.; DeSalle, R. We can't all be supermodels: the value
603 of comparative transcriptomics to the study of non-model insects. *Insect Molecular Biology*
604 **2015**, *24*, 139–154, doi:10.1111/imb.12154.

605 19. Necsulea, A.; Kaessmann, H. Evolutionary dynamics of coding and non-coding
606 transcriptomes. *Nat Rev Genet* **2014**, *15*, 734–748, doi:10.1038/nrg3802.

607 20. Alvarez, M.; Schrey, A.W.; Richards, C.L. Ten years of transcriptomics in wild populations:
608 what have we learned about their ecology and evolution? *Molecular Ecology* **2015**, *24*, 710–
609 725, doi:10.1111/mec.13055.

610 21. Todd, E.V.; Black, M.A.; Gemmell, N.J. The power and promise of RNA-seq in ecology and
611 evolution. *Molecular Ecology* **2016**, *25*, 1224–1241, doi:10.1111/mec.13526.

612 22. Arif, S.; Kittelmann, S.; McGregor, A.P. From shavenbaby to the naked valley: trichome
613 formation as a model for evolutionary developmental biology. *Evol Dev* **2015**, *17*, 120–126,
614 doi:10.1111/ede.12113.

615 23. Preger-Ben Noon, E.; Sabarís, G.; Ortiz, D.M.; Sager, J.; Liebowitz, A.; Stern, D.L.; Frankel,
616 N. Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer
617 Function. *Cell Reports* **2018**, *22*, 3021–3031, doi:10.1016/j.celrep.2018.02.073.

618 24. Rebeiz, M.; POOL, J.E.; Kassner, V.a.; AQUADRO, C.F.; Carroll, S.B. Stepwise
619 modification of a modular enhancer underlies adaptation in a Drosophila population. *Science*
620 **2009**, *326*, 1663–1667, doi:10.1126/science.1178357.

621 25. Gompel, N.; Prud'homme, B.; Wittkopp, P.J.; Kassner, V.a.; Carroll, S.B. Chance caught on
622 the wing: Cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature*
623 **2005**, *433*, 481–487, doi:10.1038/nature03235.

624 26. Manceau, M.; Domingues, V.S.; Mallarino, R.; Hoekstra, H.E. The developmental role of
625 Agouti in color pattern evolution. *Science* **2011**, *331*, 1062–1065,
626 doi:10.1126/science.1200684.

627 27. Hoekstra, H.E. Genetics, development and evolution of adaptive pigmentation in vertebrates.
628 *Heredity* **2006**, *97*, 222–234, doi:10.1038/sj.hdy.6800861.

629 28. Kratochwil, C.F.; Liang, Y.; Gerwin, J.; Woltering, J.M.; Urban, S.; Henning, F.; Machado-
630 Schiaffino, G.; Hulsey, C.D.; Meyer, A. Agouti-related peptide 2 facilitates convergent
631 evolution of stripe patterns across cichlid fish radiations. *Science* **2018**, *362*, 457–460,
632 doi:10.1126/science.aao6809.

633 29. Hagen, J.F.D.; Mendes, C.C.; Tanaka, K.M.; Gaspar, P.; Kittelmann, M.; McGregor, A.P.;
634 Nunes, M.D.S. tartan underlies the evolution of male genital morphology. *bioRxiv* **2018,**
635 **doi:**10.1101/462259.

636 30. Nagy, O.; Nuez, I.; Savisaar, R.; Peluffo, A.E.; Yassin, A.; Lang, M.; Stern, D.L.; Matute,
637 D.R.; David, J.R.; Courtier-Orgogozo, V. Correlated Evolution of Two Copulatory Organs via
638 a Single cis-Regulatory Nucleotide Change. *Current Biology* **2018**, *28*, 3450-3457.e13,
639 doi:10.1016/j.cub.2018.08.047.

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. *Nature Communications* **2018**, *9*, 4338, doi:10.1038/s41467-018-06824-8.

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, doi:10.1038/nature22074.

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, doi:10.1073/pnas.1213423109.

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 Genetics* **2017**, *13*, e1006907, doi:10.1371/journal.pgen.1006907.

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, doi:10.1186/s13059-018-1444-y.

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 Genetics* **2013**, *9*, e1003813, doi:10.1371/journal.pgen.1003813.

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 Genetics* **2016**, *12*, e1005987, doi:10.1371/journal.pgen.1005987.

38. Crocker, A.; Guan, X.-J.; Murphy, C.T.; Murthy, M. Cell-Type-Specific Transcriptome Analysis in the Drosophila Mushroom Body Reveals Memory-Related Changes in Gene Expression. *Cell Reports* **2016**, *15*, 1580–1596, doi:10.1016/j.celrep.2016.04.046.

39. Perou, C.M.; Sørlie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; Rees, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Akslen, L.A.; et al. Molecular portraits of human breast tumours. *Nature Insight Biodiverstiy* **2000**, *406*, 747–752, doi:10.1038/35021093.

40. Lübbe, A.; Schaffner, W. Tissue-specific gene expression. *Trends in Neurosciences* **1985**, *8*, 100–104, doi:10.1016/0166-2236(85)90046-3.

41. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragoczy, 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, doi:10.1126/science.1181369.

42. 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, doi:10.1016/j.cell.2012.01.010.

43. 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, doi:10.1038/nature11049.

44. 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, doi:10.1038/nature11082.

45. 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, doi:10.1126/science.1234167.

46. 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.e19, doi:10.1016/j.cell.2016.09.037.

47. Buenrostro, 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. *Nature Methods* **2013**, *10*, 1213–1218, doi:10.1038/nmeth.2688.

48. 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, doi:**10.1016/j.cell.2018.06.052.

49. 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, doi:10.1038/nature25981.

50. Lawrence, M.; Daujat, S.; Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics* **2016**, *32*, 42–56, doi:10.1016/j.tig.2015.10.007.

51. 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, doi:10.1016/j.tibs.2017.10.004.

52. Johnson, D.S.; Mortazavi, A.; Myers, R.M.; Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)* **2007**, *316*, 1497–1502, doi:10.1126/science.1141319.

53. 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 Meth* **2007**, *4*, 651–657, doi:10.1038/nmeth1068.

54. 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, doi:10.1038/nature13415.

55. 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 Research* **2011**, *21*, 227–236, doi:10.1101/gr.115519.110.

56. 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, doi:10.1126/science.1237973.

57. 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. *Molecular Cell* **2008**, *32*, 232–246, doi:10.1016/j.molcel.2008.08.022.

58. 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, doi:10.1038/nature09819.

59. 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, doi:10.1038/nature12210.

60. 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, doi:10.1080/15476286.2015.1112487.

61. 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 Research* **2011**, *21*, 203–215, doi:10.1101/gr.116657.110.

62. 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 Research* **2018**, *28*, 52–65, doi:10.1101/gr.226068.117.

63. 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 Research* **2016**, *44*, 3865–3877, doi:10.1093/nar/gkw116.

64. Bird, A.P.; Wolffe, A.P. Methylation-Induced Repression— Belts, Braces, and Chromatin. *Cell* **1999**, *99*, 451–454, doi:10.1016/S0092-8674(00)81532-9.

65. Jaenisch, R.; Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **2003**, *33 Suppl*, 245–254, doi:10.1038/ng1089.

66. Zhu, H.; Wang, G.; Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* **2016**, *17*, 551–565, doi:10.1038/nrg.2016.83.

67. Mohn, F.; Schübeler, D. Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends in Genetics* **2009**, *25*, 129–136, doi:10.1016/j.tig.2008.12.005.

68. Spitz, F.; Furlong, E.E.M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **2012**, *13*, 613–626, doi:10.1038/nrg3207.

69. 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, doi:10.1016/j.cell.2018.01.029.

70. Homem, C.C.F.; Knoblich, J.A. Drosophila neuroblasts: A model for stem cell biology. *Development (Cambridge, England)* **2012**, *139*, 4297–4310, doi:10.1242/dev.080515.

71. Karcavich, R.E. Generating neuronal diversity in the Drosophila central nervous system: a view from the ganglion mother cells. *Developmental Dynamics* **2005**, *232*, 609–616, doi:10.1002/dvdy.20273.

72. Technau, G.M.; Berger, C.; Urbach, R. Generation of cell diversity and segmental pattern in the embryonic central nervous system of Drosophila. *Developmental Dynamics* **2006**, *235*, 861–869, doi:10.1002/dvdy.20566.

73. Nogales, E.; Louder, R.K.; He, Y. Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu. Rev. Biophys.* **2017**, *46*, 59–83, doi:10.1146/annurev-biophys-070816-033751.

74. Engel, C.; Neyer, S.; Cramer, P. Distinct Mechanisms of Transcription Initiation by RNA Polymerases I and II. *Annu. Rev. Biophys.* **2018**, *47*, 425–446, doi:10.1146/annurev-biophys-070317-033058.

75. Levine, M.; Tjian, R. Transcription regulation and animal diversity. *Nature* **2003**, *424*, 147–151, doi:10.1038/nature01763.

76. Banerji, J.; Rusconi, S.; Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **1981**, *27*, 299–308.

77. 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. *Proceedings of the National Academy of Sciences* **1996**, *93*, 9322–9328, doi:10.1073/pnas.93.18.9322.

78. Arnone, M.I.; Davidson, E.H. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **1997**, *124*, 1851–1864.

79. Small, S.; Blair, A.; Levine, M. Regulation of even-skipped stripe 2 in the Drosophila embryo. *The EMBO Journal* **1992**, *11*, 4047–4057.

80. 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, doi:10.1016/0092-8674(89)90916-1.

81. Weasner, B.M.; Weasner, B.P.; Neuman, S.D.; Bashirullah, A.; Kumar, J.P. Retinal Expression of the Drosophila eyes absent Gene Is Controlled by Several Cooperatively Acting Cis-regulatory Elements. *PLoS Genetics* **2016**, *12*, 1–31, doi:10.1371/journal.pgen.1006462.

82. Kvon, E.Z.; Kazmar, T.; Stampfel, G.; Yáñez-Cuna, J.O.; Pagani, M.; Schernhuber, K.; Dickson, B.J.; Stark, A. Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **2014**, *512*, 91–95, doi:10.1038/nature13395.

83. Haenlin, M.; Cubadda, Y.; Blondeau, F.; Heitzler, P.; Lutz, Y.; Simpson, P.; Ramain, 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 & Development* **1997**, *11*, 3096–3108, doi:10.1101/gad.11.22.3096.

84. Fromental-Ramain, C.; Taquet, N.; Ramain, P. Transcriptional interactions between the pannier isoforms and the cofactor U-shaped during neural development in Drosophila. *Mechanisms of Development* **2010**, *127*, 442–457, doi:10.1016/j.mod.2010.08.002.

85. Fromental-Ramain, C.; Vanolst, L.; Delaporte, C.; Ramain, P. pannier encodes two structurally related isoforms that are differentially expressed during Drosophila development

and display distinct functions during thorax patterning. *Mechanisms of Development* **2008**, *125*, 43–57, doi:10.1016/j.mod.2007.10.008.

86. 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. *Current Biology* **2014**, *24*, 1–10, doi:10.1016/j.cub.2013.11.017.

87. Elkon, R.; Ugalde, A.P.; Agami, R. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet* **2013**, *14*, 496–506, doi:10.1038/nrg3482.

88. Mayr, C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends in Cell Biology* **2016**, *26*, 227–237, doi:10.1016/j.tcb.2015.10.012.

89. Topisirovic, I.; Svitkin, Y.V.; Sonenberg, N.; Shatkin, A.J. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdisciplinary Reviews: RNA* **2011**, *2*, 277–298, doi:10.1002/wrna.52.

90. Matera, A.G.; Wang, Z. A day in the life of the spliceosome. *Nature Reviews Molecular Cell Biology* **2014**, *15*, 108–121, doi:10.1038/nrm3742.

91. Grosso, A.R.; Gomes, A.Q.; Barbosa-Morais, N.L.; Caldeira, S.; Thorne, N.P.; Grech, G.; Lindern, M. von; Carmo-Fonseca, M. Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Research* **2008**, *36*, 4823–4832, doi:10.1093/nar/gkn463.

92. Bartel, D.P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **2004**, *116*, 281–297, doi:10.1016/S0092-8674(04)00045-5.

93. Fatica, A.; Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* **2014**, *15*, 7–21, doi:10.1038/nrg3606.

94. Gong, C.; Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **2011**, *470*, 284–288, doi:10.1038/nature09701.

95. 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, doi:10.1038/nature11661.

96. 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, doi:10.1038/nature09144.

97. Alvarez-Garcia, I.; Miska, E.A. MicroRNA functions in animal development and human disease. *Development* **2005**, *132*, 4653–4662, doi:10.1242/dev.02073.

98. 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. *Nature Communications* **2019**, *10*, 821, doi:10.1038/s41467-019-08679-z.

99. 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. *Current Biology* **2018**, *28*, 3547-3561.e9, doi:10.1016/j.cub.2018.09.026.

100. Mercer, T.R.; Mattick, J.S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300–307, doi:10.1038/nsmb.2480.

101. O'Neil, D.; Glowatz, H.; Schlumpberge, M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr. Protoc. Mol. Biol.* **2013**, 1–8, doi:10.1002/0471142727.mb0419s103.

102. 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, doi:10.1080/15476286.2018.1514236.

103. 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, doi:10.1038/nmeth.1226.

104. Pritchard, C.C.; Cheng, H.H.; Tewari, M. *MicroRNA profiling: Approaches and considerations*, 2012. *Nature Reviews Genetics*, *13* (5). http://www.targetscan.org.

105. Akhade, V.S.; Pal, D.; Kanduri, C. Long Noncoding RNA: Genome Organization and Mechanism of Action. *Advances in Experimental Medicine and Biology* **2017**, *1008*, 47–74, doi:10.1007/978-981-10-5203-3_2.

106. 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, doi:10.1038/nmeth.1315.

107. 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. *Nature Communications* **2017**, *8*, 14049, doi:10.1038/ncomms14049.

108. Svensson, V.; Vento-Tormo, R.; Teichmann, S.A. Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols* **2018**, *13*, 599–604, doi:10.1038/nprot.2017.149.

109. 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*, 21.29.1-21.29.9, doi:10.1002/0471142727.mb2129s109.

110. 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, doi:10.1038/nmeth.4396.

111. 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 Research* **2017**, *27*, 2096–2107, doi:10.1101/gr.226233.117.

112. 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, doi:10.1126/science.aab1601.

113. 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. *Nature Insight Biodiverstiy* **2015**, *523*, 486–490, doi:10.1038/nature14590.

114. 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, doi:10.1126/science.1186176.

115. Rotem, A.; Ram, O.; Shoresh, 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, doi:10.1038/nbt.3383.

116. 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, doi:10.1016/J.YMETH.2012.05.001.

117. Sati, S.; Cavalli, G. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* **2017**, *126*, 33–44, doi:10.1007/s00412-016-0593-6.

118. 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. *Nature Methods* **2015**, *12*, 71–78, doi:10.1038/nmeth.3205.

119. 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, doi:10.1016/J.CELL.2015.05.048.

120. 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, doi:10.1038/nature11082.

121. 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 Research* **2019**, *47*, 648–665, doi:10.1093/nar/gky1103.

122. 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. *Nature Insight Biodiverstiy* **2013**, *502*, 59–64, doi:10.1038/nature12593.

123. 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. *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 1827–1831, doi:10.1073/pnas.89.5.1827.

124. 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. *Nature Insight Biodiverstiy* **2008**, *452*, 215–219, doi:10.1038/nature06745.

125. 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, doi:10.1038/s41559-017-0229-0.

126. Yong, W.-S.; Hsu, F.-M.; Chen, P.-Y. Profiling genome-wide DNA methylation. *Epigenetics & chromatin* **2016**, *9*, 26, doi:10.1186/s13072-016-0075-3.

127. Verhoeven, K.J.F.; VonHoldt, B.M.; Sork, V.L. Epigenetics in ecology and evolution: what we know andÂ what we need to know. *Molecular Ecology* **2016**, *25*, 1631–1638, doi:10.1111/mec.13617.

128. 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 Research* **2013**, *23*, 2126–2135, doi:10.1101/gr.161679.113.

129. 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. *Nature Methods* **2014**, *11*, 817–820, doi:10.1038/nmeth.3035.

130. 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, doi:10.1038/nature10532.

131. 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, doi:10.1126/science.1108296.

132. Yang, J.; Su, A.I.; Li, W.-H. Gene expression evolves faster in narrowly than in broadly expressed mammalian genes. *Molecular Biology and Evolution* **2005**, *22*, 2113–2118, doi:10.1093/molbev/msi206.

133. Irie, N.; Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications* **2011**, *2*, 248, doi:10.1038/ncomms1248.

134. Domazet-Lošo, T.; Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **2010**, *468*, 815–818, doi:10.1038/nature09632.

135. Liscovitch, N.; Chechik, G. Specialization of gene expression during mouse brain development. *PLOS Computational Biology* **2013**, *9*, e1003185, doi:10.1371/journal.pcbi.1003185.

136. Penso-Dolfin, L.; Moxon, S.; Haerty, W.; Di Palma, F. The evolutionary dynamics of microRNAs in domestic mammals. *Sci Rep* **2018**, *8*, 17050, doi:10.1038/s41598-018-34243-8.

137. Hezroni, H.; Koppstein, D.; Schwartz, M.G.; Avrutin, A.; Bartel, D.P.; Ulitsky, I. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports* **2015**, *11*, 1110–1122, doi:10.1016/j.celrep.2015.04.023.

138. 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. *Frontiers in Genetics* **2014**, *5*, 43, doi:10.3389/fgene.2014.00043.

139. 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 Reports* **2014**, *9*, 2290–2303, doi:10.1016/j.celrep.2014.11.038.

140. 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. *Nature Neuroscience* **2013**, *16*, 1896–1905, doi:10.1038/nn.3554.

141. 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. *Scientific Reports* **2017**, *7*, 1343, doi:10.1038/s41598-017-01468-y.

142. 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,** **doi:**10.1126/science.aaa1975.

143. 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**, *3235*, eaan3235-eaan3235, doi:10.1126/science.aan3235.

144. Sebé-Pedrós, A.; Saudemont, B.; Chomsky, E.; Plessier, F.; Mailhé, 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.e20, doi:10.1016/j.cell.2018.05.019.

145. Fincher, C.T.; Wurtzel, O.; Hoog, T. de; Kravarik, K.M.; Reddien, P.W. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science* **2018**, *360, doi:*10.1126/science.aaq1736.

146. 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. *Molecular Biology and Evolution* **2018**, *35*, 1047–1062, doi:10.1093/molbev/msx336.

147. Davie, K.; Janssens, J.; Koldere, D.; Waegeneer, M. de; Pech, U.; Kreft, Ł.; Aibar, S.; Makhzami, S.; Christiaens, V.; Bravo González-Blas, C.; et al. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **2018, doi:**10.1016/j.cell.2018.05.057.

148. Konstantinides, N.; Kapuralin, K.; Fadil, C.; Barboza, L.; Satija, R.; Desplan, C. Phenotypic Convergence: Distinct Transcription Factors Regulate Common Terminal Features. *Cell* **2018, doi:**10.1016/j.cell.2018.05.021.

149. Croset, V.; Treiber, C.D.; Waddell, S. Cellular diversity in the Drosophila midbrain revealed by single-cell transcriptomics. *eLife* **2018**, *7, doi:*10.7554/eLife.34550.

150. Plass, M.; Solana, J.; Wolf, F.A.; Ayoub, S.; Misios, A.; Glažar, P.; Obermayer, B.; Theis, F.J.; Kocks, C.; Rajewsky, N. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science (New York, N.Y.)* **2018**, *360, doi:*10.1126/science.aaq1723.

151. Trapnell, C.; Cacchiarelli, D.; Grimsby, J.; Pokharel, P.; Li, S.; Morse, M.; Lennon, N.J.; Livak, K.J.; Mikkelsen, T.S.; Rinn, J.L. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **2014**, *32*, 381–386, doi:10.1038/nbt.2859.

152. Kolodziejczyk, A.A.; Kim, J.K.; Tsang, J.C.H.; Ilicic, T.; Henriksson, J.; Natarajan, K.N.; Tuck, A.C.; Gao, X.; Bühler, M.; Liu, P.; et al. Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell* **2015**, *17*, 471–485, doi:10.1016/j.stem.2015.09.011.

153. Aibar, S.; González-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, doi:10.1038/nmeth.4463.

154. Chan, T.E.; Stumpf, M.P.H.; Babtie, A.C. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst.* **2017**, *5*, 251-267.e3, doi:10.1016/j.cels.2017.08.014.

155. 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. *Nature Genetics* **2000**, *25*, 25–29, doi:10.1038/75556.

156. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* **2019**, *47*, D330-D338, doi:10.1093/nar/gky1055.

157. 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. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 15545–15550, doi:10.1073/pnas.0506580102.

158. Mootha, V.K.; Lindgren, C.M.; Eriksson, K.-F.; Subramanian, A.; Sihag, S.; Lehar, J.; Puigserver, P.; Carlsson, E.; Ridderstråle, M.; Laurila, E.; et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **2003**, *34*, 267–273, doi:10.1038/ng1180.

159. Goeman, J.J.; Bühlmann, P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics (Oxford, England)* **2007**, *23*, 980–987, doi:10.1093/bioinformatics/btm051.

160. Nam, D.; Kim, S.-Y. Gene-set approach for expression pattern analysis. *Briefings in bioinformatics* **2008**, *9*, 189–197, doi:10.1093/bib/bbn001.

161. 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 Research* **2011**, *39*, W316-22, doi:10.1093/nar/gkr483.

162. 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. *Frontiers in plant science* **2017**, *8*, 1427, doi:10.3389/fpls.2017.01427.

163. 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 Genomics* **2019**, *20*, 21, doi:10.1186/s12864-018-5356-8.

164. 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, doi:*10.1093/gigascience/giy027.

165. Rockman, M.V.; Kruglyak, L. Genetics of global gene expression. *Nat Rev Genet* **2006**, *7*, 862–872, doi:10.1038/nrg1964.

166. Gilad, Y.; Rifkin, S.a.; Pritchard, J.K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics* **2008**, *24*, 408–415, doi:10.1016/j.tig.2008.06.001.

167. Jia, Z.; Xu, S. Mapping quantitative trait loci for expression abundance. *Genetics* **2007**, *176*, 611–623, doi:10.1534/genetics.106.065599.

168. 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. *Nature Genetics* **2007**, *39*, 1202–1207, doi:10.1038/ng2109.

169. Knight, J.C. Allele-specific gene expression uncovered. *Trends in Genetics* **2004**, *20*, 113–116, doi:10.1016/j.tig.2004.01.001.

170. Coolon, J.D.; Wittkopp, P.J. cis- and trans -Regulation in Drosophila Interspecific Hybrids **2015**, *305*, 37–57, doi:10.1002/9781118552872.ch3.

171. Korir, P.K.; Seoighe, C. Inference of allele-specific expression from RNA-seq data. *Methods Mol Biol* **2014**, *1112*, 49–69, doi:10.1007/978-1-62703-773-0_4.

172. 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, doi:10.1534/genetics.117.201459.

173. 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, doi:10.1534/genetics.109.105957.

174. 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. *Molecular Biology and Evolution* **2012**, *29*, 1521–1532, doi:10.1093/molbev/msr318.

175. 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. *Molecular Ecology* **2010**, *19*, 212–227, doi:10.1111/j.1365-294X.2010.04472.x.

176. 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 Research* **2010**, *20*, 816–825, doi:10.1101/gr.102491.109.

177. Tirosh, I.; Reikhav, S.; Levy, A.a.; Barkai, N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science (New York, N.Y.)* **2009**, *324*, 659–662, doi:10.1126/science.1169766.

178. Zhang, X.; Borevitz, J.O. Global analysis of allele-specific expression in Arabidopsis thaliana. *Genetics* **2009**, *182*, 943–954, doi:10.1534/genetics.109.103499.

179. Lemmon, Z.H.; Bukowski, R.; Sun, Q.; Doebley, J.F. The Role of cis Regulatory Evolution in Maize Domestication. *PLoS Genetics* **2014**, *10, doi:*10.1371/journal.pgen.1004745.

180. 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, doi:10.1186/gb-2008-9-8-r125.

181. Lappalainen, T.; Sammeth, M.; Friedländer, M.R.; Hoen, P.a.C. 't; Monlong, J.; Rivas, M.A.; Gonzàlez-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, doi:10.1038/nature12531.

182. Emilsson, V.; Thorleifsson, G.; Zhang, B.; Leonardson, A.S.; Zink, F.; Zhu, J.; Carlson, S.; Helgason, A.; Walters, G.B.; Gunnarsdottir, S.; et al. Genetics of gene expression and its effect on disease. *Nature* **2008**, *452*, 423–428, doi:10.1038/nature06758.

183. 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, doi:10.1038/nature06258.

184. 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, doi:10.1016/j.cell.2015.07.048.

185. Degner, J.F.; Pai, A.a.; Pique-Regi, R.; Veyrieras, J.-B.; Gaffney, D.J.; Pickrell, J.K.; Leon, S. de; 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, doi:10.1038/nature10808.

186. 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, doi:10.1126/science.1242429.

187. 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 Meth* **2015**, *12*, 458–464, doi:10.1038/nmeth.3326.

188. Arnold, P.; Schöler, A.; Pachkov, M.; Balwierz, P.J.; Jørgensen, H.; Stadler, M.B.; van Nimwegen, E.; Schübeler, D. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Research* **2013**, *23*, 60–73, doi:10.1101/gr.142661.112.

189. 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. *Nature Genetics* **2014**, *46*, 685–692, doi:10.1038/ng.3009.

190. Li, Y.I.; 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, doi:10.1126/science.aad9417.

191. Monlong, J.; Calvo, M.; Ferreira, P.G.; Guigó, R. Identification of genetic variants associated with alternative splicing using sQTLseekeR. *Nature Communications* **2014**, *5*, 4698, doi:10.1038/ncomms5698.

192. Han, J.; Xiong, J.; Wang, D.; Fu, X.-D. Pre-mRNA splicing: where and when in the nucleus. *Trends in Cell Biology* **2011**, *21*, 336–343, doi:10.1016/j.tcb.2011.03.003.

193. Shukla, S.; Kavak, E.; Gregory, M.; Imashimizu, M.; Shutinoski, B.; Kashlev, M.; Oberdoerffer, P.; Sandberg, R.; Oberdoerffer, S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **2011**, *479*, 74–79, doi:10.1038/nature10442.

194. 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. *Nature Communications* **2018**, *9*, 804, doi:10.1038/s41467-018-03209-9.

195. 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, doi:10.1038/ng1990.

196. 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, doi:10.7554/eLife.05538.

197. 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 Research* **2006**, *16*, 995–1004, doi:10.1101/gr.5217506.

198. He, F.; Arce, A.L.; Schmitz, G.; Koornneef, M.; Novikova, P.; Beyer, A.; Meaux, J. de. The Footprint of Polygenic Adaptation on Stress-Responsive Cis-Regulatory Divergence in the Arabidopsis Genus. *Molecular Biology and Evolution* **2016**, *33*, 2088–2101, doi:10.1093/molbev/msw096.

199. Meaux, J. de; Pop, A.; Mitchell-Olds, T. Cis-regulatory evolution of chalcone-synthase expression in the genus Arabidopsis. *Genetics* **2006**, *174*, 2181–2202, doi:10.1534/genetics.106.064543.

200. Wittkopp, P.J.; Haerum, B.K.; Clark, A.G. Regulatory changes underlying expression differences within and between Drosophila species. *Nature Genetics* **2008**, *40*, 346–350, doi:10.1038/ng.77.

201. Glaser-Schmitt, A.; Parsch, J. Functional characterization of adaptive variation within a cis-regulatory element influencing Drosophila melanogaster growth. *PLoS Biol* **2018**, *16*, e2004538, doi:10.1371/journal.pbio.2004538.

202. Lind, M.I.; Spagopoulou, F. Evolutionary consequences of epigenetic inheritance. *Heredity* **2018**, *121*, 205–209, doi:10.1038/s41437-018-0113-y.

203. Hu, J.; Barrett, R.D.H. Epigenetics in natural animal populations. *Journal of Evolutionary Biology* **2017**, *30*, 1612–1632, doi:10.1111/jeb.13130.

204. Barros-Silva, D.; Marques, C.J.; Henrique, R.; Jerónimo, C. Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. *Genes (Basel)* **2018**, *9, doi:*10.3390/genes9090429.

205. Nagalingam, K.; Lorenc, M.T.; Manoli, S.; Cameron, S.L.; Clarke, A.R.; Dudley, K.J. Chromatin immunoprecipitation (ChIP) method for non-model fruit flies (Diptera: Tephritidae) and evidence of histone modifications. *PLoS ONE* **2018**, *13*, e0194420, doi:10.1371/journal.pone.0194420.

206. Woltering, J.M.; Noordermeer, D.; Leleu, M.; Duboule, D. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS Biol* **2014**, *12*, e1001773, doi:10.1371/journal.pbio.1001773.