Cis-acting variation is common across regulatory layers but is often buffered during embryonic development

Swann Floc'Hlay, Emily S Wong, Bingqing Zhao, Rebecca R Viales, Morgane Thomas-Chollier, Denis Thieffry, David A Garfield, Eileen E M Furlong

▶ **To cite this version:**

Swann Floc'Hlay, Emily S Wong, Bingqing Zhao, Rebecca R Viales, Morgane Thomas-Chollier, et al.. Cis-acting variation is common across regulatory layers but is often buffered during embryonic development. Genome Research, 2021, 31, pp.211 - 224. 10.1101/gr.266338.120 . hal-03352083

HAL Id: hal-03352083
https://hal.science/hal-03352083

Submitted on 22 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

Cis-acting variation is common across regulatory layers but is often buffered during embryonic development

Swann Floc'hlay,[1,6,7] Emily S. Wong,[2,3,6,7] Bingqing Zhao,[4,6,7] Rebecca R. Viales,[4] Morgane Thomas-Chollier,[1,5] Denis Thieffry,[1] David A. Garfield,[4] and Eileen E.M. Furlong[4]

[1]*Institut de Biologie de l'ENS (IBENS), École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France;* [2]*Molecular, Structural and Computational Biology Division, Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales 2010, Australia;* [3]*School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Kensington, New South Wales 2052, Australia;* [4]*European Molecular Biology Laboratory (EMBL), Genome Biology Unit, D-69117 Heidelberg, Germany;* [5]*Institut Universitaire de France (IUF), 75005 Paris, France*

Precise patterns of gene expression are driven by interactions between transcription factors, regulatory DNA sequences, and chromatin. How DNA mutations affecting any one of these regulatory "layers" are buffered or propagated to gene expression remains unclear. To address this, we quantified allele-specific changes in chromatin accessibility, histone modifications, and gene expression in F1 embryos generated from eight *Drosophila* crosses at three embryonic stages, yielding a comprehensive data set of 240 samples spanning multiple regulatory layers. Genetic variation (allelic imbalance) impacts gene expression more frequently than chromatin features, with metabolic and environmental response genes being most often affected. Allelic imbalance in *cis*-regulatory elements (enhancers) is common and highly heritable, yet its functional impact does not generally propagate to gene expression. When it does, genetic variation impacts RNA levels through two alternative mechanisms involving either H3K4me3 or chromatin accessibility and H3K27ac. Changes in RNA are more predictive of variation in H3K4me3 than vice versa, suggesting a role for H3K4me3 downstream from transcription. The impact of a substantial proportion of genetic variation is consistent across embryonic stages, with 50% of allelic imbalanced features at one stage being also imbalanced at subsequent developmental stages. Crucially, buffering, as well as the magnitude and evolutionary impact of genetic variants, is influenced by regulatory complexity (i.e., number of enhancers regulating a gene), with transcription factors being most robust to *cis*-acting, but most influenced by *trans*-acting, variation.

[Supplemental material is available for this article.]

The development of a multicellular organism requires tight regulation of gene expression in both space and time to ensure that reproducible phenotypes are obtained across individuals and environmental conditions. DNA regulatory elements (e.g., promoters and enhancers) are essential to this process by integrating regulatory information from sequence-specific transcription factors (TFs), RNA polymerase II (Pol II), and other regulatory proteins to drive specific spatiotemporal patterns of expression during development. But although gene expression patterns are typically quite precise, the DNA regulatory elements that control these patterns are replete with genetic variation (mutations), which can impact transcriptional regulation at multiple levels, including TF binding (Kasowski et al. 2010; Spivakov et al. 2012; Behera et al. 2018), chromatin state (Waszak et al. 2015), transcriptional start site (TSS) usage (Schor et al. 2017), gene expression levels (Garfield et al. 2013; Battle et al. 2015; Cannavò et al. 2017), and transcript isoform diversity (Cannavò et al. 2017).

Although regulatory mutations can have large effects, many behave effectively neutrally, making predictions of the functional impact of genetic variants extremely challenging. Part of the difficulty comes from a general lack of knowledge about which regions of noncoding DNA have regulatory (not just biochemical) function. Another major challenge is the inherent robustness of gene regulatory networks. At least within a laboratory context, sections of regulatory DNA can be removed with little apparent impact on phenotype or fitness (Ahituv et al. 2007). Similarly, divergent regulatory sequences from different species can be experimentally swapped with few detectable changes in gene expression across species (Borok et al. 2010). Developmental systems have built-in redundancy that can "buffer" the effects of regulatory mutations, for example, through compensation by other regulatory elements with partially overlapping activities (Hong et al. 2008; Frankel et al. 2010; Cannavò et al. 2016).

The complex relationship between DNA sequence and regulatory output further complicates our understanding of how mutations can impact gene regulation. For example, mutations affecting TF binding motifs can have a large impact on chromatin accessibility, Pol II occupancy, histone modifications, and gene expression (Kircher et al. 2019). But in some contexts/tissues, TF binding is driven by collective processes that can include protein–protein and protein–DNA interactions, such that mutations affecting a single TF motif may not substantially affect TF recruitment (Junion et al. 2012; Doitsidou et al. 2013; Uhl et al. 2016;

[6]These authors contributed equally to this work.
[7]These authors are listed alphabetically.
Corresponding authors: dagarfield@gmail.com; furlong@embl.de
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.266338.120. Freely available online through the *Genome Research* Open Access option.

© 2021 Floc'hlay et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

Khoueiry et al. 2017). Moreover, many sequence variants affecting TF occupancy in vivo lie outside the TF's cognate motif and are likely owing to variation affecting the binding of co-occurring factors (Kasowski et al. 2010; Zheng et al. 2010; Reddy et al. 2012) or an overall change in DNA shape (Levo et al. 2015). To make matters more complex, enhancer output is not a strict function of all factors that occupy an enhancer; enhancers often contain binding sites for multiple factors with redundant input, and in some cases, different combinations of TFs can produce the same expression output (Brown et al. 2007; Zinzen et al. 2009; Khoueiry et al. 2017). Even in cases in which an enhancer's activity is abolished by mutations, gene expression may not be affected as genes often have many enhancers with partially overlapping activity (Hong et al. 2008; Frankel et al. 2010; Cannavò et al. 2016). With a few exceptions (Bullaughey 2011), this complex genotype-to-phenotype relationship cannot be modelled using regulatory sequence information alone but rather must be evaluated empirically (Khoueiry et al. 2017).

Allelic-specific data provide a unique opportunity to study the molecular mechanisms of *cis*-acting variation and have uncovered multiple regulatory processes through which *cis*-acting variation impacts transcriptional control (Kilpinen et al. 2013; Chen et al. 2016). F1 crosses of inbred strains provide an elegant method to determine the contribution of both *cis* and *trans* variation (Wittkopp et al. 2004; Tirosh et al. 2009; Goncalves et al. 2012; Wong et al. 2017). By using a multiline F1 design, we sought to better understand how natural sequence variation impacts different steps of gene regulation during embryonic development. We collected *Drosophila* F1 hybrid embryos from eight crosses and quantified allele-specific changes in TF occupancy (using open chromatin [ATAC-seq] as a proxy), enhancer and promoter activity (using H3K27ac or H3K4me3 and H3K27ac ChIP-seq as proxies, respectively), and gene expression (RNA-seq). By treating genetic variation affecting each of these regulatory layers as a perturbation to gene regulation, we could uncover functional relationships between different regulatory layers during embryonic development, as well as their impact on gene expression.

Results

Quantifying gene expression and regulatory element activity in hybrid embryos

To generate genetically diverse samples suitable for allele-specific analyses, we mated eight genetically distinct inbred lines from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) collection (Mackay et al. 2012) to females from a common isogenic maternal line (Fig. 1A). The resulting half-sibling F1 panel contains an average of 567,412 SNPs per cross and a total of 1,455,988 unique SNPs covering a range of minor allele-frequencies and levels of conservation (phyloP scores) (Supplemental Fig. S1A; Supplemental Table S1).

Embryos were collected at three stages of embryogenesis: 2–4 h after egg laying, consisting primarily of pregastrulation, unspecified embryos (mainly stage 5); 6–8 h (mainly stage 11), when major lineages within the three germ-layers are specified; and 10–12 h (mainly stage 13), during terminal differentiation of tissues (Fig. 1A). For each developmental stage and F1 cross, we performed RNA-seq (gene expression), ATAC-seq (open chromatin), and iChIP for H3K27ac (marking active enhancers and promoters) and H3K4me3 (active promoters) (Buenrostro et al. 2013; Lara-Astiaso et al. 2014) from the same collection of embryos (four measurements × three stages × eight genotypes = 96 samples). In addition, we collected samples from the parental lines of one cross, forming a parent/offspring trio that allowed us to partition genetic differences between the parents into *cis* and *trans* (Wittkopp et al. 2004), defined here as genetic variation that affects the linked alleles of features on the same chromosome (*cis*) versus variation that affects both alleles on any chromosome (*trans*). All measurements were made in replicates, giving a total of 240 samples (192 F1 samples [96 × two replicates] + 48 parental [four measurement × three stages × two genotypes × two replicates]). Read counts were highly correlated between biological replicates, with median correlation coefficients of 0.98 for RNA, ATAC, and histone data (Methods) (Supplemental Fig. S1B). As expected, correlations were reduced between corresponding samples across genotypes, reflecting the functional impact of genetic variation (Supplemental Fig. S1C), and were reduced even further across time points, reflecting dynamic changes in gene expression during development.

To define noncoding features, ATAC-seq and ChIP-seq reads from each cross were mapped to each parental genome independently and the significant peaks merged into a combined peak set used for all subsequent analyses. In total, we identified 11,211 genes with detectable expression, 31,963 ATAC peaks, 19,769 H3K27ac peaks, and 6648 H3K4me3 peaks active at one or more stages (Supplemental Table S2). Of these, 93.9%, 95.8%, 95.2%, and 96.9%, respectively, contained at least one SNP that distinguishes maternal and paternal haplotypes in at least one line. The *CG12402* locus, a predicted ubiquitin-protein transferase, illustrates the dynamics of the data, transitioning from low to high expression from 2–4 h to 10–12 h (Fig. 1A), accompanied by quantitative changes in chromatin accessibility and, to a lesser extent, histone modifications in its promoter region.

To examine the regulatory relationships between these different signals, we divided noncoding features into promoter-proximal (<±500 bp of an annotated TSS or H3K4me3 peak [to capture unannotated promoters]) or -distal (putative enhancer; >±500 bp from a TSS) elements. At promoter-proximal regions, all signals show the expected enrichment and distribution around the TSS (Fig. 1B, proximal), showing the quality of the data. The ATAC-seq signal is highest directly at the promoter, representing occupancy of the basal transcriptional machinery, whereas H3K27ac and H3K4me3 signals are highest at the +1 nucleosome, reflecting the predominantly unidirectional nature of *Drosophila* promoters (Core et al. 2012; Mikhaylichenko et al. 2018). Although all three regulatory signals (ATAC-seq, H3K27ac, and H3K4me3) are highly correlated at promoters of actively transcribed genes (i.e., with RNA-seq signal), many promoter-proximal peaks (3907) marked by H3K4me3 with ATAC signal and/or H3K27ac show no detectable RNA signal (Fig. 1C, left upset plot). Many of these regions (approximately 850) correspond to known noncoding RNAs not captured by poly(A)$^+$ RNA-seq libraries, suggesting the presence of many additional unannotated ncRNAs.

The majority of H3K27ac (62.8%) and ATAC peaks (63.9%) are distal to an annotated promoter. Distal ATAC peaks lacking H3K27ac signal (58% of the total) have a strong enrichment for Polycomb and Su(Hw) ChIP signal (Supplemental Table S3), suggesting that they represent repressed enhancers or other types of regulatory elements (e.g., insulators). For the remaining 9007 distal elements, H3K27ac signal is bimodally distributed around the ATAC-seq peak (Fig. 1B), suggestive of active enhancers. Although most H3K27ac peaks (60%) overlap ATAC peaks, many do not (Fig. 1C, right). This latter set often have ATAC signal below

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

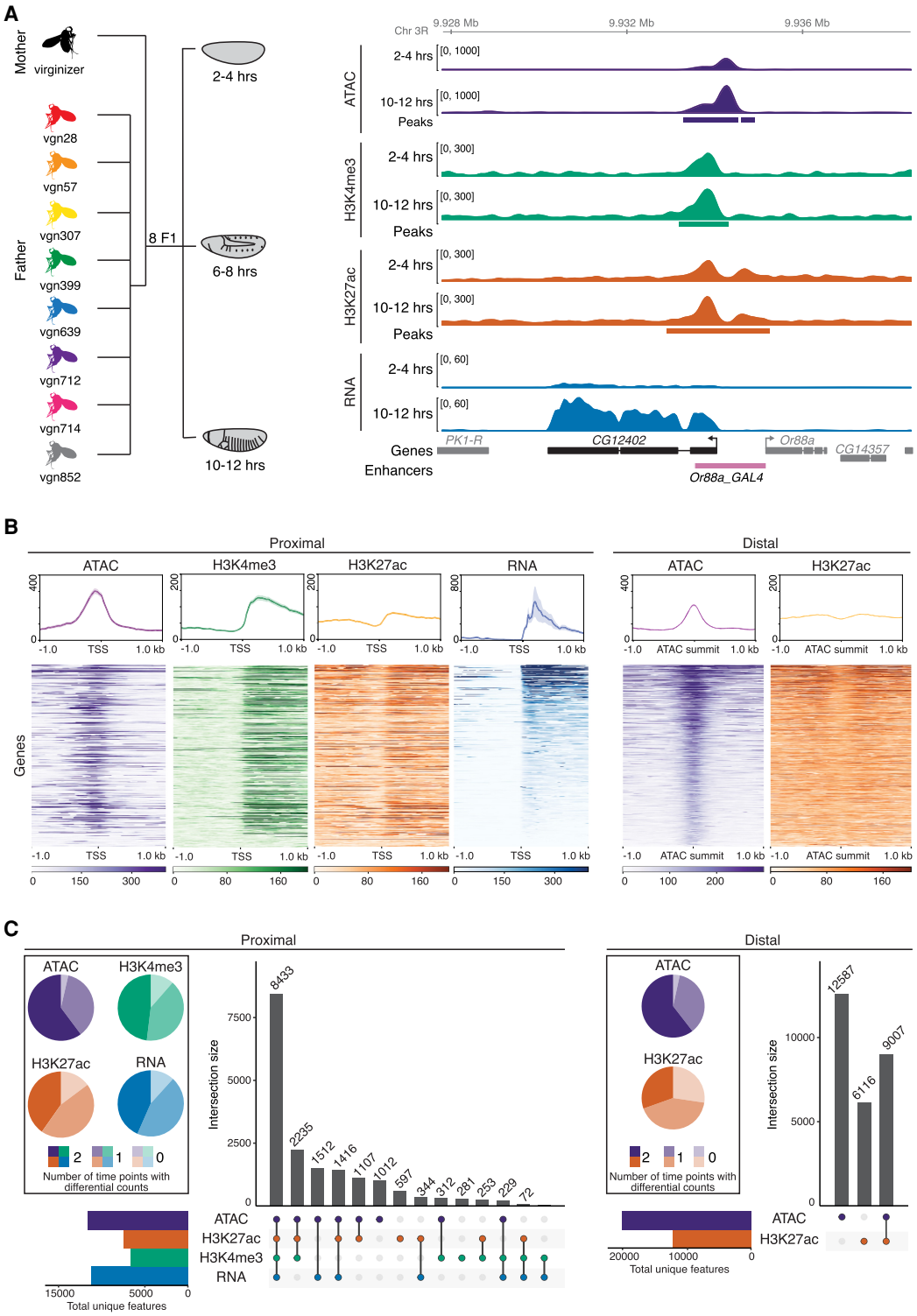

Figure 1. Quantifying gene expression and regulatory element activity in F1 embryos. (*A*, *left*) Experimental design and data structure. RNA-seq, ATAC-seq, and iChIP of H3K4me3 and H3K27ac were performed on embryos at three developmental stages from eight F1 hybrids with a common maternal line. (*Right*) Genome browser overview for the *CG12402* locus showing all data for 2–4 h and 10–12 h for the genotype vgn28. *Bottom* track shows curated enhancers (Kvon et al. 2014). (*B*, *top*) Density plots for representative cross (vgn × DGRP-639) at 6–8 h showing read count signal for TSS proximal and distal regions. Plots centered at TSS for promoter-proximal (*left*) and ATAC summits for distal (*right*) regions; shaded regions indicate 95% confidence intervals. (*Bottom*) Heatmaps showing quantitative signal of the same data as above, where rows were sorted by mean RNA-seq (proximal) or ATAC-seq (distal) signal. (*C*) *Upset* plots show colocalization of signal for proximal and distal regions for all four data types over all genotypes and stages. Regions common between data types (filled circle) are joined by a vertical bar. Horizontal bar plots indicate the number of unique genes/features. Pie charts indicate proportion of features with statistically different total read counts between time points (color indicates the number of times [0/1/2] the feature is differentially expressed).

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

our threshold for peak calling (Supplemental Fig. S1D) and are enriched in a range of factors including Elav (Oktaba et al. 2015) and H3K27me3, suggesting that they represent a mixture of regulatory features, including enhancers with active and repressed states.

To quantify dynamic changes in each of the four regulatory layers across embryonic development, we pooled all F1 samples within a time point to formally test for changes in read counts across time (treating the data as 16 replicates per time point). Both proximal and distal sites, gene expression (RNA-seq), and noncoding elements (based on ATAC-seq and chromatin signatures) show similar dynamics, with the majority (72%–96%) of features showing statistically significant changes in total counts between developmental time points across all *F1* lines (Methods) (Fig. 1C, pie charts).

Taken together, these features show both the quality and richness of the data and their usefulness to further annotate the regulatory landscape of the *Drosophila* genome at these important stages of embryogenesis.

Allele-specific variation is common across genotypes and regulatory layers

To quantify the impact of *cis*-acting genetic variation, we compared the number of reads mapping to the maternal and paternal chromosomes in each F1 cross, using informative SNPs to assign reads to parent of origin (Methods) and an empirical Bayes framework to formally test for imbalance in the ratio of maternal:paternal reads at each locus per line and time point combination (Supplemental Fig. S2A). As expected, most genes and regulatory features had allelic ratios centered at 50:50 across autosomes (Fig. 2A; Supplemental Fig. S2B), with a slight elevation in the magnitude of allelic imbalance (AI) at distal sites (Supplemental Fig. S3A). RNA allelic ratios were also concordant with the direction of change of embryonic eQTL (Supplemental Fig. S3B), previously quantified in the same paternal lines at the same stages of embryogenesis (Cannavò et al. 2017), further verifying our approach and the quality of the data. The early embryonic time point (2–4 h) is an expected exception to this balanced maternal/paternal ratio owing to maternally deposited transcripts and the presence of unfertilized eggs (Supplemental Fig. S2B). We

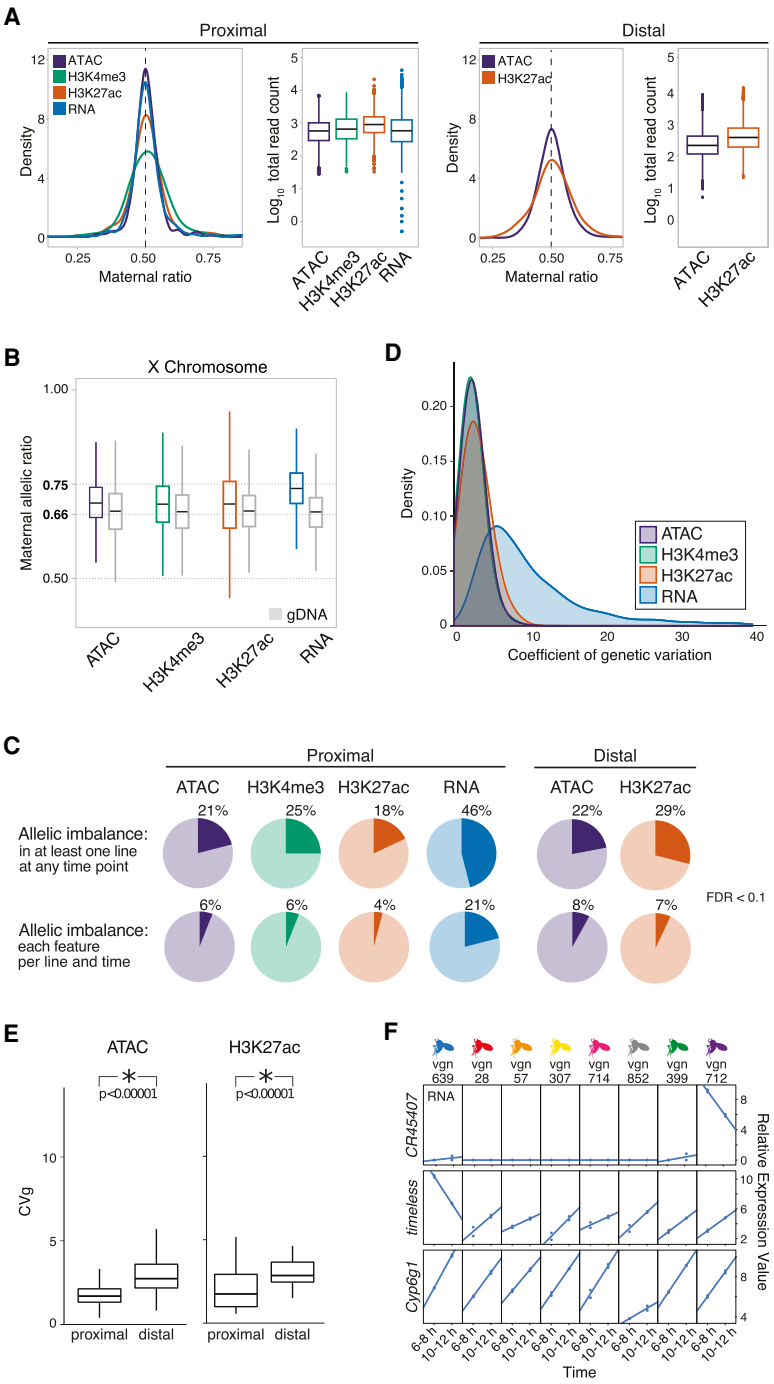

Figure 2. Allelic imbalance (AI) is common across regulatory data types. (*A*) Density plot of allelic count distribution for representative cross (vgn × DGRP-028) at 10–12 h and matching box plots showing total read count abundance ($\log_{10}$) in autosomes at TSS proximal (*left*) and distal (*right*) regions. Density plots for all crosses shown in Supplemental Figure S2B. (*B*) Box plot shows the distribution of maternal allelic ratios on the X Chromosome compared with genomic DNA (gray). (*C*) Pie charts show AI genes/features at promoter-proximal (*left*; TSS ±500 bp) and promoter-distal (*right*; 500–1500 bp ± from TSS) regions for all data types (FDR < 0.1). (*Top*) AI events in at least one F1 line at any time point. (*Bottom*) AI events detected in all eight F1 lines in all time points, on a per line and time basis. (*D*) Smoothed histograms showing distribution of coefficients of genetic variation for all features with significant between-line variances within each regulatory layer. (*E*) Box plots show the distribution of coefficient of genetic variation (CVg; *y*-axis) for chromatin accessibility (*left*) and H3K27ac signal (*right*), for promoter-proximal and promoter-distal sites. Genetic influences are more pronounced at distal elements in ATAC and H3K27ac. (*F*) Three examples of individual lines having distinct expression variation on specific genes. Relative expression values are typically larger for RNA than noncoding features, an effect that often results from one or two lines as shown here.

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

therefore removed the 2–4 h time point RNA samples from all allele-specific analyses.

To evaluate sex ratios in the embryo pools and to set a reference point for evaluating AI and dosage compensation on the X Chromosome (Lucchesi and Kuroda 2015), we sequenced the genomic DNA (gDNA) of each cross. This confirmed that our embryonic pools were relatively sex balanced, with the expected X Chromosome allelic ratio of approximately 0.66 observed for gDNA (Fig. 2B). Sex-chromosome dosage compensation in *Drosophila* is achieved by a two-fold up-regulation in gene expression of X Chromosome genes in XY males (Georgiev et al. 2011). Consistent with this, we observed a maternal:paternal ratio of 0.74 for RNA, closely matching the expected 0.75 ratio (Methods) (Fig. 2B). A proposed hypothesis for this twofold up-regulation in gene expression on the male X Chromosome is that it is due to a twofold increase in the loading of polymerase at the corresponding promoter (Conrad et al. 2012). However, the maternal:paternal ratio that we observe for chromatin data does not fully reflect this up-regulation: For both chromatin accessibility and histone modifications, the observed ratios at X Chromosome sites are more similar to the observed genomic ratio of 0.66 than to the expected 0.75 ratio under full dosage compensation (H3K27ac = 0.688, H3K4me3 = 0.692) (Fig. 2B). This indicates that dosage compensation in *Drosophila* does not involve a linear increase in chromatin accessibility on the male X Chromosome (although we do observe a slight increase in accessibility; ATAC = 0.693) (Fig. 2B; Urban et al. 2017; Pal et al. 2019). Regardless of its cause, we used the empirically observed average ratio for X Chromosome features for each data type to form the null hypothesis in subsequent beta-binomial tests for AI.

Overall, AI is common, with 46% of genes and between 18% and 25% of noncoding chromatin features showing statistically significant AI in at least one line at one or more time point (FDR < 0.1) (Fig. 2C). The magnitude of AI is generally evenly distributed across SNPs with a range of minor allelic frequencies. However, highly imbalanced peaks show a strong enrichment for extremely rare SNPs found uniquely in the maternal line relative to the DGRP panel (χ^2 test, $P < 2.2 \times 10^{-16}$) (Supplemental Fig. S3C), highlighting the disproportionate impact of rare mutations on expression phenotypes (Cannavò et al. 2017).

AI is more frequently observed for RNA than for other regulatory layers (Fig. 2C), in contrast to mammals (Goncalves et al. 2012; Wong et al. 2017). Additionally, *Drosophila* promoter-proximal sequences appear to evolve more rapidly than distal elements (putative enhancers); proximal elements are slightly more polymorphic (pair-wise differences [pi] = 0.132 vs. 0.129; Wilcoxon test, $P = 1 \times 10^{-10}$) and evolve faster (phyloP = 0.514 vs. 0.560; Wilcoxon test, $P < 2.2 \times 10^{-16}$) (Supplemental Table S4). Despite this, distal peaks of open chromatin and H3K27ac show larger (Tukey's ASD, $P < 1 \times 10^{-4}$) and slightly more frequent (χ^2 test, $P < 2.2 \times 10^{-16}$) AI than their proximal counterparts (Supplemental Fig. S3A).

To understand how allelic imbalance relates to heritable variation at the total count level, we took advantage of the fact that our measured F1 lines share a common maternal line but have unrelated, genetically diverse paternal lines. As a result, differences among F1s are expected to be proportional to heritability, or the degree to which phenotypic variation can be explained by genetic factors (Lynch and Walsh 1998). Expressed as percentage deviation from the mean phenotype (coefficient of genetic variation), the impact of genetic variation on chromatin features is relatively modest, with the average peak varying by ~5%–10% of the mean phenotype among crosses (Fig. 2D). The magnitude of heritability genetic variation is generally higher at distal, compared with proximal, regulatory elements ($P < 1 \times 10^{-5}$; Methods) (Fig. 2E), consistent with the greater magnitude of AI at distal sites. In contrast to chromatin features, the magnitude of effects is generally higher for RNA, with an average coefficient of genetic variation of ~9% and a tail extending to ~40%. In many of these cases, high coefficients are driven by one or a few lines showing highly divergent patterns of expression (Fig. 2F), suggesting the presence of large effect, likely *cis*-acting mutations.

AI is pronounced in metabolism and environmental response genes

Categorical enrichment for AI (Methods) identifies more extensive imbalance for genes associated with fast-evolving, *Drosophila*-specific genes (Mi et al. 2003; Turner et al. 2008) and metabolic genes, whereas TFs and their associated regulatory elements are depleted (Supplemental Fig. S4; Supplemental Table S5), consistent with our previous eQTL study (Cannavò et al. 2017). Selection may play a role in these AI differences; regulatory regions in the vicinity of TFs show reduced nucleotide diversity (pi, rank biserial correlation = −0.052; $P < 1 \times 10^{-4}$) and harbor more low-frequency SNPs (rank biserial correlation = −0.173; $P = 2.8 \times 10^{-3}$) (Supplemental Table S5) compared with background. However, this difference in AI may also be explained by different sensitivities between gene categories to mutations, a point we explore below.

For most gene categories, AI is equally likely to favor the maternal or the paternal allele. However, immunity and insecticide resistance genes show a clear paternal bias (Supplemental Fig. S5A; Supplemental Table S6). *Cyp6g1*, for example, is not expressed in embryos of our laboratory-derived maternal line (Supplemental Fig. S5B) but is strongly up-regulated in every measured paternal haplotype from the wild, and its expression contributes to DDT resistance in multiple *Drosophila* species (Supplemental Fig. S5B; Daborn et al. 2001; Battlay et al. 2016). Highly imbalanced genes like *Cyp6g1* (Supplemental Table S6) often overlap genes whose expression varies extensively among lines ($P < 1 \times 10^{-6}$) (Supplemental Fig. S5C) and have high levels of heritability, suggesting a close link between *cis*-acting variation and selection to changing environments.

The impact of *cis*-acting genetic variation is largely consistent across development

Gene expression is highly dynamic across development (Arbeitman et al. 2002) and is largely driven by dynamic changes in enhancer usage (Wilczyński and Furlong 2010; Reddington et al. 2020). However, the extent to which this dynamism shapes the impact of regulatory variation is not well understood. We thus examined how allelic ratios at individual loci changed during embryogenesis.

Overall, we observed considerable constancy of AI between embryonic time points; imbalanced features at one time point have a ~50% chance of being imbalanced in the subsequent time point (Fig. 3A; Supplemental Fig. S6A). To further quantify this, we constructed a series of linear models comparing the effect sizes of genetics (genotype/line) versus developmental stages (time), and the interaction between the two (GxT), for both total counts and allelic ratios.

For total counts, developmental time was the greatest contributor to variation across all data types (Fig. 3B, upper panel), consistent with the clear time-specific clustering by principal

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

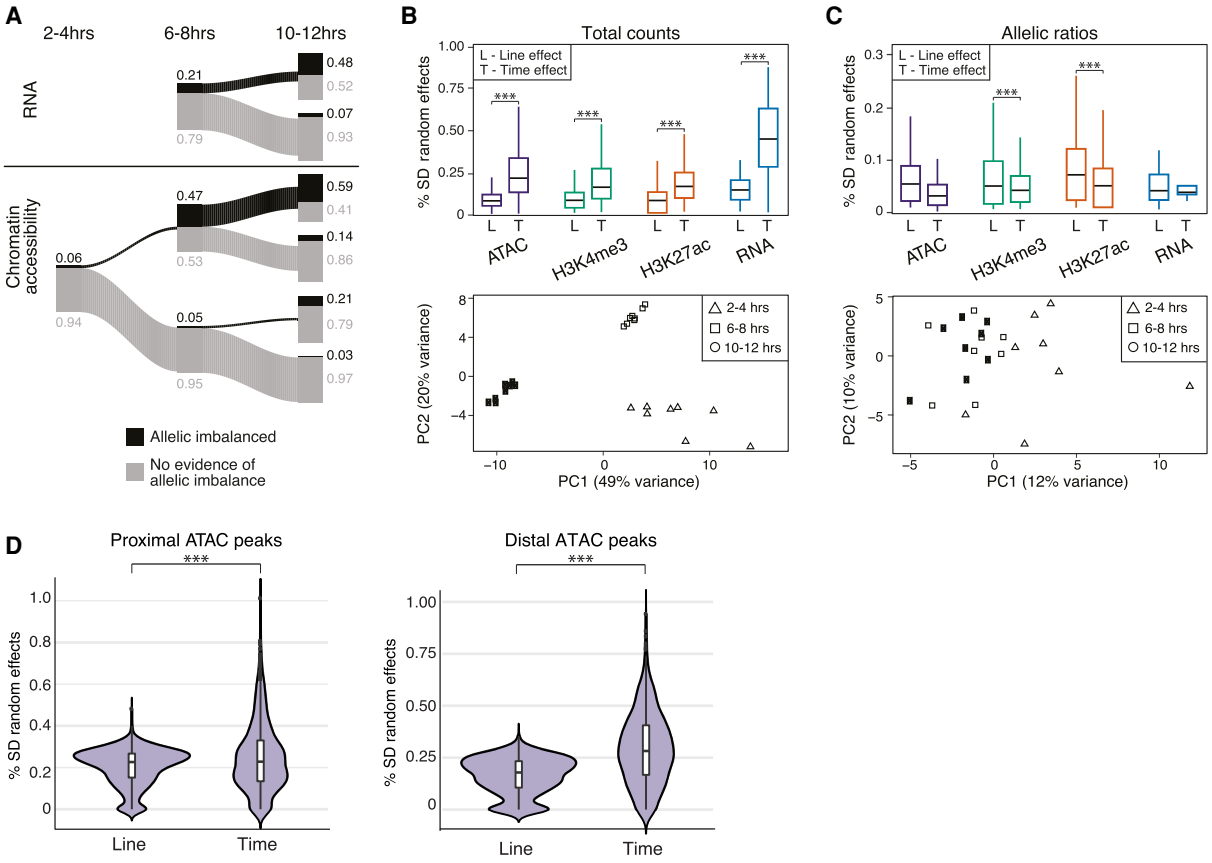

Figure 3. AI is generally not predictive of developmental time. (*A*) The relationship of AI across time for RNA (*top*) and chromatin accessibility (*bottom*). Proportions of AI and non-AI features shown in black and gray, respectively, and represented by the thickness of line (exact numbers indicated). Data for RNA at 2–4 h are not included owing to presence of maternal transcripts. (*B*, *top*) Box plots show distribution of time (T) and line (L) effect sizes obtained from mixed linear models for total counts. (*Bottom*) Principal component analysis (PCA) of gene expression for total counts. (*C*, *top*) Box plots showing the distribution of time (T) and line (L) effect sizes obtained from mixed linear models for allelic ratios. (*Bottom*) PCA of gene expression for allelic ratios. (*D*) Results from mixed linear models examining the effect of developmental time versus line (genotype) between proximal and distal ATAC-seq peaks for total count data. Distal peaks show a larger time effect compared with genotype effect (Mann–Whitney *U* test, $P < 2.2 \times 10^{-61}$). This is only slightly evident for promoter-proximal peaks (Mann–Whitney *U* test, $P < 8.5 \times 10^{-5}$).

component analysis (PCA; Methods) (shown for RNA in Fig. 3B, lower panel). Time effects are more pronounced at distal compared with proximal ATAC peaks (Fig. 3D), reflecting both the constitutive accessibility of many promoters and the dynamic usage of enhancers during development (Wilczyński and Furlong 2010; Reddington et al. 2020). Interaction effects (GxT) occur frequently and are particularly common for gene expression, making up ~30% of all analyzed models (Supplemental Table S7), consistent with our previous analyses at the total count level (Cannavò et al. 2017).

In contrast, the impact of time is significantly reduced compared with genetic effects for allelic ratios (Fig 3C, upper panel). Correspondingly, there is a lack of time point–specific clustering of allelic ratios in PCA (Fig 3C, lower panel), although there are some examples of allelic ratios that change over time in a coordinated manner between regulatory layers (Supplemental Fig. S6B). Unlike total counts, there is little evidence for interaction effects (Supplemental Table S7), consistent with the rarity of gene × environment effects reported for AI (Moyerbrailean et al. 2016; Knowles et al. 2017).

In summary, allelic effects are often larger at distal compared with promoter regions, with effects at both regions being largely consistent across developmental time points. In contrast, total counts vary markedly between embryonic time points, with interactions between genotype and developmental stage (GxT) being common.

Genetic variation affects gene expression through chromatin by two different mechanisms

Although highly correlated, the causal relationships between chromatin accessibility, histone modifications, and gene expression remain unclear. To assess this, we used allelic ratios as a perturbation to different regulatory layers and modelled the paths by which genetic variation impacts regulatory phenotypes. Like total counts, allelic ratios are highly correlated among regulatory layers (Fig. 4A; Supplemental Fig. S7), and in all cases, we could reject the null hypothesis of independence (all *P*-values $< 4.2 \times 10^{-17}$). Co-occurrence of statistically significant imbalance (intersection-union test, FDR < 0.1) is pronounced for chromatin features, in particular H3K4me3 and H3K27ac, at promoter regions, with a log-odds greater than 2.0 (Methods) (Fig. 4B). For chromatin accessibility and H3K27ac, the co-occurrence of AI is more pronounced at promoters than enhancers (distal) (Fig. 4B), despite AI being

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

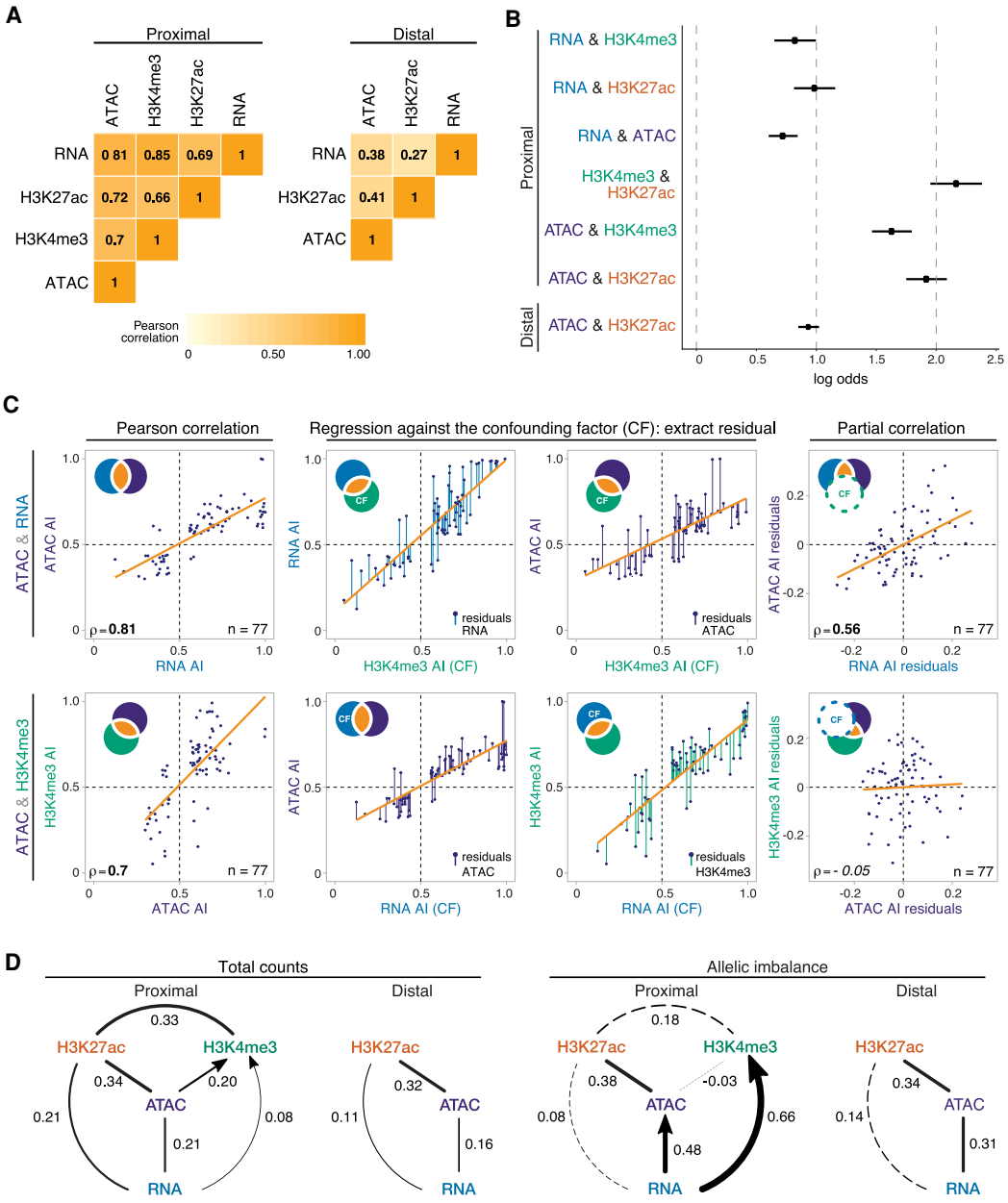

Figure 4. AI is propagated through regulatory layers via different epigenetic paths. (*A*) Heatmaps show Pearson correlation coefficient of allelic ratios between each data type for promoter-proximal (*left*) and promoter-distal (*right*) regions. Data restricted to 6–8 h and 10–12 h and features/genes whose allelic ratio exceeds 0.5 ± 0.06. (*B*) Log odds (intersection-union test) of co-occurrence of AI between two regulatory layers: 6- to 8-h and 10- to 12-h data shown; bars (from dots), 95% confidence intervals. (*C*) Stepwise example of partial correlation analysis of allelic ratios between chromatin accessibility and gene expression (ATAC&RNA; *top*), and promoter-proximal chromatin accessibility and H3K4me3 (ATAC&H3K4me3; *bottom*). Venn diagram schematics (*top left*) illustrate the variance of each variable and its shared proportion (orange), measured by linear regression (orange lines). (*Left* panels) Pearson's correlations for the two comparisons are significant. (*Middle*) Regression of each initial variable against a third, confounding variable (H3K4me3, *upper* row; RNA, *lower* row). Residuals of the initial variables (colored lines) represent the nonoverlapping part of the circle of the same color in the schematic. (*Right*) Correlations of the residuals, excluding shared variance by the confounding factor (dashed circle in schematic). This resulting partial correlation is significant in the *top*, and not in the *bottom*, examples. (*D*) Partial correlation and directional dependency regression for total counts (*left*) and allelic ratios (*right*). Significant partial correlations (solid lines) suggest direct dependencies among regulatory layers. For each significant edge ($P < 0.01$), copula regression was used to assign directionality (arrows; delta > 0.01). Line thickness indicates the value of partial correlations; dashed lines indicate nonsignificance. Promoter-proximal and promoter-distal results shown separately.

more frequent ($P < 2.2 \times 10^{-16}$) (Fig. 2C) and being of greater magnitude (Supplemental Fig. S3A) at distal sites. This suggests that H3K27ac and chromatin accessibility are more functionally coupled at promoters compared with enhancers, perhaps reflecting the fact that not all active enhancers seem to require H3K27ac (Bonn et al. 2012; Pradeepa et al. 2016).

To identify potentially causal relationships across regulatory layers, we used partial correlation to identify independent,

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

pairwise correlations between multiple covarying variables beyond their global correlations, after thresholding on allelic ratios to remove features/genes with low information content (Methods) (Fig. 4C; Supplemental Fig. S8A; Lasserre et al. 2013; Pai et al. 2015). For total count data, our results closely mirror those of Lasserre et al. (2013) in $CD4^+$ and IMR-90 cells, including finding a clear relationship between gene expression levels and the total abundance of H3K27ac that is independent, at least at a statistical level, of the correlation between expression and H3K4me3 (i.e., there is little independent correlation between changes in RNA and changes in H3K4me3 total counts) (Fig. 4D, left). We also observed a statistically significant relationship between open chromatin and gene expression, confirming that although RNA levels are influenced by post-transcriptional processes, genetic variation at *cis*-regulatory elements contributes directly to differences in gene expression among individuals.

This relationship is even more pronounced for allelic ratios, which show a clear link between open chromatin and gene expression at both proximal and distal elements (Fig. 4D, right). H3K27ac and open chromatin are also significantly correlated at promoters, although we see little evidence for a direct relationship between H3K27ac and gene expression itself (Fig. 4D, right). The latter is a marked difference from what is observed with total count data and suggests that although H3K27ac at promoters is highly correlated with, and even predictive of (Karlic et al. 2010), gene expression, they may not be mechanistically directly linked. In contrast, allelic ratios for promoter-proximal H3K4me3 show strong evidence of a direct correlation with gene expression that is independent of allelic differences in chromatin accessibility or H3K27ac (Fig. 4D, right). Taken together, this analysis suggests two independent pathways by which segregating mutations influence gene expression: one affecting open chromatin and promoter-proximal H3K27ac and the other influencing H3K4me3.

To explore these relationships further, we analyzed each edge identified by partial correlations using copula directional dependence analysis (Kim et al. 2008; Lee and Kim 2019), a statistical approach based on copula regression that evaluates the directionality of pairwise relationships while allowing for nonlinearities (Methods). For TSS-proximal regions, our analysis placed RNA upstream of both H3K4me3 and open chromatin (Fig. 4D, right, arrow). Although counterintuitive at first glance, this suggests that gene expression is relatively robust to variation in H3K4me3, whereas conversely, variation in RNA is more predictive of H3K4me3 signal. This may reflect buffering processes but is also consistent with the hypothesis that H3K4me3 is not functionally required for transcription but is rather deposited as a consequence and may be involved in post-transcriptional events (Howe et al. 2017). Similarly, allele-specific variation in RNA better explains variation in chromatin accessibility compared with the reverse; that is, not all variation in open chromatin leads to a corresponding change in gene expression (Fig. 4D, right).

In summary, by measuring informative dependencies on the impact of *cis*-acting genetic variation, we identified multiple epigenetic pathways affecting transcription. Specifically, genetic variation acts to change gene expression levels via the interplay between at least two different promoter-proximal paths: open chromatin and H3K27ac, or H3K4me3. Moreover, the flow of information suggests that gene expression is often buffered against *cis*-acting mutations (presumably affecting TF binding) at associated regulatory elements, although we cannot rule out post-transcriptional processes that may also help to buffer allelic ratios.

Regulatory buffering varies depending on gene function and local chromatin architecture

Genes often differ in the complexity of their regulatory landscapes. Metabolic genes, for example, typically have relatively simple and compact regulatory landscapes with few enhancers that are located close to the gene's promoter (Zabidi et al. 2015; Corrales et al. 2017). TFs, in contrast, have many enhancers often with partially overlapping spatial activity ("shadow enhancers"), located at varying distances from the gene's promoter (Spitz and Furlong 2012; Long et al. 2016), which may help to buffer TFs against mutations in regulatory DNA (Xiong et al. 2002; Cretekos et al. 2008; Montavon et al. 2011; Cannavò et al. 2016; Lu and Rogan 2018). We evaluated this hypothesis in two ways. First, we assessed the extent to which AI in the expression of different gene categories is independent of, or buffered from, imbalance in their associated regulatory elements using conditional probabilities. Among all comparisons, the expression of ancient genes (conserved bilaterian processes) and of genes coding for TFs, transmembrane proteins, and signaling components is most robust to imbalance in their regulatory regions, whereas genes involved in metabolism have high sensitivity (Fig. 5A, cf. blue and orange).

Second, we directly assessed the relationship between AI and regulatory complexity (ATAC-seq peak number within a gene's ± 1.5-kb TSS regulatory domain). Imbalanced genes have fewer associated ATAC peaks genome-wide (Kruskal–Wallis, $P = 1.1 \times 10^{-16}$) (Fig. 5B), a trend that is most pronounced for single-peak genes, which have significantly more AI than genes associated with multiple regulatory elements (Mann–Whitney U test, $P = 6.4 \times 10^{-6}$). Furthermore, genes associated with previously characterized, partially redundant "shadow enhancers" (Cannavò et al. 2016) have a modest reduction in the frequency of AI compared with genes without ($x^2 = 5.3$, $P = 0.02$) (Fig. 5C). However, AI at multiple enhancers in the vicinity of a gene can have a cumulative influence on gene expression, as genes with multiple regulatory elements are more likely to be imbalanced when multiple associated peaks show unbalanced allelic ratios (Fig. 5D).

In summary, the degree to which a gene's expression is influenced by noncoding genetic variation in its regulatory elements is influenced by the gene's regulatory complexity, with more regulatory elements providing a degree of buffering against genetic perturbations.

Trans-acting variation influences the heritability of gene expression

Regulatory differences between individuals may also be influenced by *trans*-acting genetic variation. To assess the influence of *trans*-acting variation, we measured the same regulatory features in embryos from our maternal line (vgn) and one paternal line (DGRP-399), forming a trio of embryos from both parental lines (F0) and their F1 embryos. This allowed us to quantify *trans* effects by comparing the difference between parental lines (from both *cis*- and *trans*-acting variants) to allelic ratios (*cis* only) measured in the F1 (Landry et al. 2005; Tirosh et al. 2009; Goncalves et al. 2012; Wong et al. 2017) using a maximum likelihood framework that classified genes and features as *cis*, *trans*, *cistrans*, or *conserved* (Wong et al. 2017).

Among noncoding chromatin features, *cis*-acting effects are more common than *trans* (59% vs. 41%; $P < 2.2 \times 10^{-16}$, χ^2; Methods) (Fig. 6A; Supplemental Table S8). This enrichment is particularly pronounced for histone modifications, with nearly twice as many *cis* influenced peaks compared to *trans* (Fig. 6A;

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

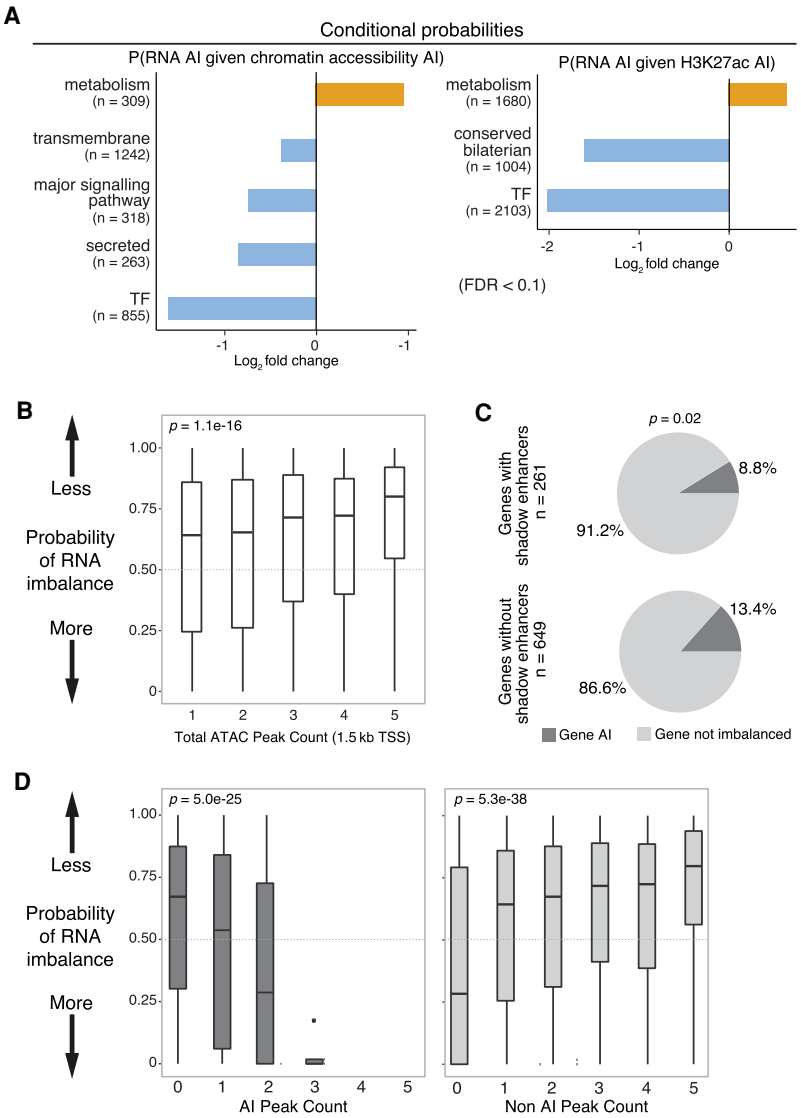

Figure 5. Regulatory buffering varies across gene categories and with local chromatin structure. (*A*) Conditional probability of AI in gene expression given AI in associated chromatin (*left*) and H3K27ac (*right*) peaks across gene categories. *X*-axis shows $\log_2$ fold change; background is genome-wide expectation. Gene categories enriched (orange) or depleted (blue) for imbalance are indicated (FDR > 0.1, Fisher's exact test). n = number of genes per category combined across all lines and both 6- to 8-h and 10- to 12-h samples. (*B*) Box plots show probabilities of AI in gene expression based on numbers of neighboring ATAC peaks (TSS < 1.5 kb). Genes with more ATAC peaks are less likely to show AI. (*C*) Pie charts show the proportion of genes with AI in RNA associated to ATAC-seq peaks overlapping known partially redundant/shadow enhancers (*top*) or not (*bottom*). Genes associated with shadow enhancers are less likely to have AI (determined using a beta-binomial model; x^2=5.3, P = 0.02). (*D*) Box plots show the probability of AI in gene expression (*P*-value from a beta-binomial model; *y*-axis) as a function of the number of ATAC-seq peaks with AI (*left*) or not imbalanced (*right*).

Supplemental Fig. S9A). For both open chromatin and H3K27ac, *cis* effects are slightly more common at promoters than enhancers (ATAC = 0.32 vs. 0.29, H2K27ac = 0.30 vs. 0.27; $P < 1 \times 10^{-4}$).

Gene expression, in contrast, is more strongly influenced by *trans*-acting genetic variation (55% *trans* vs. 45% *cis*; $P = 0.0073$, χ^2) (Fig. 6A). Moreover, a higher fraction of *cistrans* genes have more *trans*, compared with *cis*, variation (*trans* proportions 0.67 vs. 0.53; $P = 2.77 \times 10^{-5}$) (Supplemental Fig. S9A), whereas *cis* and *trans* are balanced for *cistrans* classified noncoding features.

Previous studies suggest that *trans*-influenced features generally show nonadditive inheritance (Lemos et al. 2008; Meiklejohn et al. 2014; Wong et al. 2017) and are thus less likely to be directly influenced by natural selection (Lynch and Walsh 1998). Our data suggest that open chromatin features, whether influenced by *cis* or *trans*, have primarily additive inheritance (Supplemental Fig. S9B), consistent with the finding that most variation affecting TF binding is inherited additively (Wong et al. 2017). In contrast, for gene expression, an additive model could be rejected for 32% of genes, with *trans* influenced genes departing from an additive model more frequently than *cis* (24% vs. 2%; χ^2, $P < 1 \times 10^{-4}$) (Fig. 6B). *Trans* effects are most common for genes associated with complex regulatory landscapes (Supplemental Table S9), and correspondingly, genes showing nonadditive inheritance have more ATAC peaks (2.19 peaks per gene vs. 1.82; Wilcoxon test, $P = 1.4 \times 10^{-3}$). This suggests that in addition to buffering genes from *cis*-regulatory variation, complex regulatory landscapes can influence patterns of heritability, with downstream consequences for how selection can act on gene expression phenotypes (Supplemental Fig. S9).

Discussion

In this study, we generated an extensive F1 data set to better understand the functional impact of genetic variation in regulatory DNA on embryonic gene expression and to shed light on how these effects are propagated or buffered through different regulatory layers. Our analysis revealed several new insights into the impact of regulatory mutations on transcriptional phenotypes.

First, although *cis*-acting genetic variation is common in development, its effects are not equally distributed across the genome. Allelic variation both is more frequent and has greater magnitude at distal regulatory elements (putative enhancers) compared with promoters, despite genetic variation itself being more common at promoters. This may in part be owing to differences in the relative importance of sequence content at promoters and enhancers: Many promoters, particularly for broadly expressed genes, have high tolerance to genetic variation (Schor et al. 2017). But despite having a greater magnitude, AI at distal elements is less likely to be propagated to other regulatory layers (Fig. 3), suggesting that enhancer mutations are often effectively buffered, a hypothesis that fits well with the observed robustness of gene expression to

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

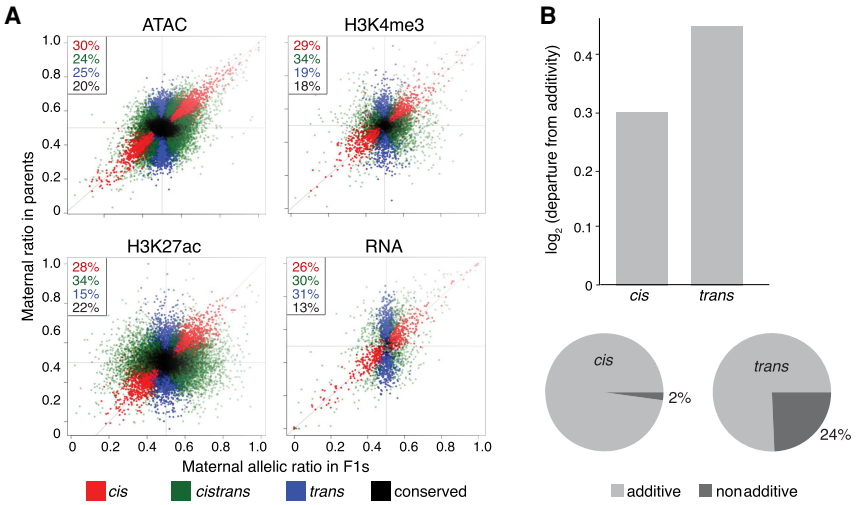

Figure 6. Chromatin features are more heritable than gene expression. (*A*) Scatter plots of F1 allelic ratio (*x*-axis) against the maternal/paternal ratio observed in (normalized) parental, total count libraries (*y*-axis). Genes/features along the diagonal are exclusively influenced by *cis*-acting variation, whereas vertically distributed genes/features show exclusively *trans* influences. Colors indicate maximum likelihood classification into *cis*, *trans*, and *cistrans* (a mixture of *cis* and *trans*) or conserved genes/features. (*B*, *top*) Bar plots showing the magnitude of deviation from additivity (parental mean) for *cis* versus *trans* (BIC ≥ 2) features. (*Bottom*) Pie charts showing fraction of additive and nonadditive genes for *cis* (*left*) and *trans* (*right*) classes.

deletions that remove distal regulatory elements (Hong et al. 2008; Cannavò et al. 2016).

Second, although all data types (open chromatin, histone modifications, RNA levels) are highly correlated, their explanatory values (potential causal relationships) as revealed by partial correlation analysis are not equal. By using *cis*-acting variation as perturbations to development, we observed a strong, likely direct, relationship between variants affecting open chromatin (TF binding) at proximal and distal sites to a degree that was not observed at the total count level. We also uncovered a strong, potentially causal, link between allelic imbalance in H3K4me3 signal and AI in the expression of the corresponding genes, which was largely independent of imbalance in H3K27ac signal. Our copula analysis placed H3K4me3 downstream from RNA (Fig. 4D), suggesting that although AI at the RNA level is predictive of AI in H3K4me3, mutations impacting H3K4me3 often do not directly impact RNA. This placement of RNA upstream of H3K4me3, inferred from our analysis of the functional impact of genetic variation, is supported by recent genetic ablation studies showing that RNA transcription does not require H3K4me3 (Clouaire et al. 2012, 2014; Margaritis et al. 2012) and is consistent with suggestions that H3K4me3 is deposited as a consequence of transcription and may be required in more downstream post-transcriptional events (Howe et al. 2017).

Third, the impact of *cis*-acting variation on gene expression is influenced by regulatory complexity, with genes having more regulatory elements being less likely to show AI (Fig. 5). This may reflect selection against variation in regulatory elements associated with these genes, and indeed, we observe less AI in regulatory elements associated with developmental regulators, extending previous findings (Cannavò et al. 2016). But even accounting for reduced overall AI, TFs and other genes with complex regulation show more independence from AI in associated regulatory layers. Although we cannot rule out a role for allele-specific post-transcriptional processes (Pai et al. 2012; Sun et al. 2018), these results suggest an active buffering process resulting from the presence of multiple regulatory inputs (Lu and Rogan 2018; Waymack et al. 2020). That said, as the number of regulatory elements with AI near a gene increases, so does the probability that the gene will show AI, suggesting that such buffering is not absolute.

Finally, *trans*-acting variation is more common for RNA than other regulatory layers, particularly for genes with complex regulatory landscapes such as TFs. This later observation, likely owing to the buffering effects of complex *cis* regulatory landscapes (Scholes et al. 2019; Waymack et al. 2020), has potentially counterintuitive evolutionary consequences: Although reduced in genetic variation overall, predominantly *trans*-influenced genes are more likely to show nonadditive, and thus less selectable, patterns of inheritance. As a result, *trans*-acting variation in genes such as TFs may remain in populations even as negative selection and buffering act to reduce the influence of *cis*-acting mutations. Why *trans*-acting variation is more common for RNA is not immediately clear but may reflect the accumulation of variation impacting both transcriptional and post-transcriptional processes (Liu et al. 2019).

In summary, allelic variation in chromatin accessibility and histone modifications at regulatory elements is prevalent and capable of propagating across regulatory layers. The extent of this information flow, or propagation, depends on the type of regulatory element and appears mitigated at developmental regulators.

Methods

Fly husbandry, crosses, and embryo collection

F1 hybrid embryos were generated by crossing males from eight genetically distinct inbred lines from the DGRP collection (Mackay et al. 2012) to females from a common maternal "virginizer" line. The virginizer line contains a heat-shock-inducible proapoptotic gene (*hid*) on the Y Chromosome (Starz-Gaiano et al. 2001) of a laboratory reference strain (w^{1118}) that kills all male embryos after a 37°C heat-shock.

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

RNA-seq, ATAC-seq, and iChIP

For three developmental stages (2–4 h, 6–8 h, and 10–12 h after egg laying), we performed RNA-seq, ATAC-seq, and iChIP for H3K27ac and H3K4me3 for pooled embryos of each F1 strain. All experiments were performed in biological replicates from independent embryo collections. iChIP experiments were performed as previously described (Lara-Astiaso et al. 2014). Detailed library information can be found in the Supplemental Methods, pages 13 through 15.

Sequencing read processing

Strain-specific genomes and annotations were constructed by inserting genetic variants and indels into the *Drosophila* dm3 assembly (version 5 from FlyBase) and translating the reference annotations (r5.57) to the personalized genomes using pslMap (Zhu et al. 2007). ATAC-seq and ChIP-seq reads were mapped using BWA (Li and Durbin 2010), whereas RNA-seq reads were mapped using STAR (Dobin et al. 2013) followed by quality filtering and the assignment of reads to parent of origin using SNP overlaps (for more details, see page 16 in the Supplemental Methods).

Demarcation of distal peaks

We evaluated the Spearman's correlations of allelic ratio between ATAC-seq and histone mark peaks to gene expression based on peaks at increasing distances from the genes' TSS up to a distance of 10 kb. We defined a distance of 1.5 kb from the TSS to be a conservative demarcation between proximal and distal peaks based on the presence of reasonable correlations between genes and associated peaks at this distance.

Test for allele-specific imbalance

Because of the extensive maternally deposited transcripts still present at 2–4 h, we excluded the RNA-seq data from this time point from all downstream allele-specific analysis to avoid potential confounding effects in AI measurements. To test for AI, an empirical Bayesian statistical framework was used to test the null hypothesis for differences in read counts between F1 alleles for each feature of each data set (RNA-seq, ATAC-seq, H3K4me3, H3K27ac) (see page 21 of the Supplemental Methods).

Allele-specific changes across lines and developmental time

A linear mixed-effects model, in which random effect components were incorporated, was used to estimate variability between pools of individuals, time points and lines:

$$y_f^{d,s,r,t} = \mu_f + \delta_f^t + \omega_f^s + (\delta\omega)_f^t \quad \omega_f^s \sim \mathrm{N}(0, \sigma_f^2)$$

μ_f is the intercept term, δ_f^t is a random effect term denoting time, ω_f^s is a random effect based on strain, and $(\delta\omega)_f^t$ is a interaction term for time by strain.

To infer the significance of time- or strain-dependent allele bias, we restricted the values that the parameters can take. Library size differences were corrected for at the allele-combined count level using the TMM method in "edgeR" (Robinson et al. 2010) before analysis. Not all features contained enough information for statistical testing; analyses were limited to features with at least six samples in each of the three time points in at least four genetic strains.

Allele-specific changes across regulatory layers

Intersection-union tests were used to examine the pairwise co-occurrence of AI in overlapping genes/features, limited to autosomes, based on rejecting the null hypothesis if a significant outcome with respect to the feature compared at the same time point exists for both data types (Berger and Jason 1996).

To infer pairwise relationships between regulatory data types while reducing indirect relations, partial correlation analysis was performed using "GeneNet" (Opgen-Rhein and Strimmer 2007) for both allelic ratios and total count data. Directional dependence modeling was performed in a regression framework using copulas (Lee and Kim 2019) to infer the flow of information for significant pairwise relationships in partial correlation analyses (see page 26 of the Supplemental Methods).

Conditional probabilities for the probability of AI given imbalance in a different regulatory data type were calculated by the following definition:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

where A and B are the probabilities of AI in each data type.

Cis/*trans* analysis

For one F1 line (vgn × 399) and its parental lines, maximum likelihood estimation (MLE) was used to compare parental and offspring ratios simultaneously to determine whether gene expression, chromatin accessibility, or H3K4me3 and H3K27ac enrichments are influenced by *cis*-acting, *trans*-acting, conserved, or *cistrans*-acting effects by modeling read counts. For parents, the total count data were modeled using negative binomial distributions whereas allelic differences in F1s were modeled using beta-binomial distributions (Supplemental Methods). We constrained parameter estimation for each model based on four different regulatory scenarios and derived maximum likelihood values for each feature (shown in Fig. 6A; Supplemental Fig. S9C). In subsequent analyses, we limited analyses to features that showed a BIC difference equal or greater than two.

Measuring additive versus nonadditive heritability

Additive inheritance implies that the F1 signal is equal to the midpoint (average) of the two parents. Nonadditive inheritance in this analysis was thus determined by testing for departure of the F1 from the parental midpoint using DESeq2 (Love et al. 2014).

Data access

All raw data generated in this study have been submitted to the EMBL-EBI ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) under accession numbers E-MTAB-8877 (gDNA), E-MTAB-8878 (RNA-seq), E-MTAB-8879 (ATAC-seq), and E-MTAB-8880 (ChIP-seq H3K4me3, H3K27ac). Processed data, including total counts, allelic ratios, *cis*/*trans* estimates, estimated per-feature heritability, mappability filters, and parental genotype files, can all be downloaded from http://furlonglab.embl.de/data.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank members of the Furlong laboratory for discussions and comments, in particular Olga Sigalova, Adam Rabinowitz, Marijn van Jaarsveld, and Matteo Perino. We thank Ido Amit and Ronnie Blecher, Weizmann Institute, for iChIP training and advice. This work was technically supported by the EMBL

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

Genomics Core Facility and public resources FlyBase, BDGP, and RedFly. The work was financially supported by a PhD grant from the French Ministry of Research (MNRT) and Fondation de la Recherche Médicale (FRM) to S.F., an Australian Research Council Discovery Early Career Researcher Award (DE16010 0755) to E.S.W., and the European Research Council (ERC advanced grant) agreement 787611 (DeCRyPT) to E.E.M.F.

Author contributions: E.E.M.F., D.A.G., and B.Z. conceived the project. B.Z. developed the ATAC-seq; R.R.V., the iChIP protocol for *Drosophila* embryos. B.Z. and R.R.V. generated all data with help from D.A.G. D.A.G., S.F., and E.S.W. performed data analysis. D.A.G. and S.F. performed mapping bias analysis and allelic and total count data processing. S.F. performed partial correlation analysis, and E.S.W. performed the statistical modelling for allelic imbalance and time by line, copula, and conditional dependence analyses. D.A.G. and E.S.W. performed *cis/trans* analysis. D.A.G. performed analysis of heritability and evolutionary variation and gene category analyses. E.E.M.F., D.A.G., E.S.W., B.Z., and S.F. wrote the manuscript with input from all authors. M.T.-C., D.T., and E.E.M.F. supervised the computational analyses performed by S.F.

References

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol* **5:** e234. doi:10.1371/journal.pbio.0050234

Arbeitman MN, Furlong EE, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297:** 2270–2275. doi:10.1126/science.1072152

Battlay P, Schmidt JM, Fournier-Level A, Robin C. 2016. Genomic and transcriptomic associations identify a new insecticide resistance phenotype for the selective sweep at the *Cyp6g1* locus of *Drosophila melanogaster*. *G3 (Bethesda)* **6:** 2573–2581. doi:10.1534/g3.116.031054

Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, Gilad Y. 2015. Genomic variation. impact of regulatory variation from RNA to protein. *Science* **347:** 664–667. doi:10.1126/science.1260793

Behera V, Evans P, Face CJ, Hamagami N, Sankaranarayanan L, Keller CA, Giardine B, Tan K, Hardison RC, Shi J, et al. 2018. Exploiting genetic variation to uncover rules of transcription factor binding and chromatin accessibility. *Nat Commun* **9:** 782. doi:10.1038/s41467-018-03082-6

Berger R, Jason C. 1996. Bioequivalence trials, intersection–union tests and equivalence confidence sets. *Stat Sci* **11:** 283–319. doi:10.1214/ss/1032280304

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EE. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44:** 148–156. doi:10.1038/ng.1064

Borok MJ, Tran DA, Ho MC, Drewell RA. 2010. Dissecting the regulatory switches of development: lessons from enhancer evolution in *Drosophila*. *Development* **137:** 5–13. doi:10.1242/dev.036160

Brown CD, Johnson DS, Sidow A. 2007. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science* **317:** 1557–1560. doi:10.1126/science.1145893

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10:** 1213–1218. doi:10.1038/nmeth.2688

Bullaughey K. 2011. Changes in selective effects over time facilitate turnover of enhancer sequences. *Genetics* **187:** 567–582. doi:10.1534/genetics.110.121590

Cannavò E, Khoueiry P, Garfield DA, Geeleher P, Zichner T, Gustafson EH, Ciglar L, Korbel JO, Furlong EE. 2016. Shadow enhancers are pervasive features of developmental regulatory networks. *Curr Biol* **26:** 38–51. doi:10.1016/j.cub.2015.11.034

Cannavò E, Koelling N, Harnett D, Garfield D, Casale FP, Ciglar L, Gustafson HE, Viales RR, Marco-Ferreres R, Degner JF, et al. 2017. Genetic variants regulating expression levels and isoform diversity during embryogenesis. *Nature* **541:** 402–406. doi:10.1038/nature20802

Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, Yan Y, Kundu K, Ecker S, et al. 2016. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167:** 1398–1414.e24. doi:10.1016/j.cell.2016.10.026

Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, Lee JH, Skalnik D, Bird A. 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev* **26:** 1714–1728. doi:10.1101/gad.194209.112

Clouaire T, Webb S, Bird A. 2014. Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome Biol* **15:** 451. doi:10.1186/s13059-014-0451-x

Conrad T, Cavalli FM, Vaquerizas JM, Luscombe NM, Akhtar A. 2012. *Drosophila* dosage compensation involves enhanced Pol II recruitment to male X-linked promoters. *Science* **337:** 742–746. doi:10.1126/science.1221428

Core LJ, Waterfall JJ, Gilchrist DA, Fargo DC, Kwak H, Adelman K, Lis JT. 2012. Defining the status of RNA polymerase at promoters. *Cell Rep* **2:** 1025–1035. doi:10.1016/j.celrep.2012.08.034

Corrales M, Rosado A, Cortini R, van Arensbergen J, van Steensel B, Filion GJ. 2017. Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome Res* **27:** 1153–1161. doi:10.1101/gr.211433.116

Cretekos CJ, Wang Y, Green ED, Martin JF, Rasweiler J, Behringer RR. 2008. Regulatory divergence modifies limb length between mammals. *Genes Dev* **22:** 141–151. doi:10.1101/gad.1620408

Daborn P, Boundy S, Yen J, Pittendrigh B, ffrench-Constant R. 2001. DDT resistance in *Drosophila* correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Mol Genet Genomics* **266:** 556–563. doi:10.1007/s004380100531

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Doitsidou M, Flames N, Topalidou I, Abe N, Felton T, Remesal L, Popovitchenko T, Mann R, Chalfie M, Hobert O. 2013. A combinatorial regulatory signature controls terminal differentiation of the dopaminergic nervous system in *C. elegans*. *Genes Dev* **27:** 1391–1405. doi:10.1101/gad.217224.113

Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466:** 490–493. doi:10.1038/nature09158

Garfield DA, Runcie DE, Babbitt CC, Haygood R, Nielsen WJ, Wray GA. 2013. The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLoS Biol* **11:** e1001696. doi:10.1371/journal.pbio.1001696

Georgiev P, Chlamydas S, Akhtar A. 2011. *Drosophila* dosage compensation: Males are from Mars, females are from Venus. *Fly (Austin)* **5:** 147–154. doi:10.4161/fly.5.2.14934

Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. 2012. Extensive compensatory *cis-trans* regulation in the evolution of mouse gene expression. *Genome Res* **22:** 2376–2384. doi:10.1101/gr.142281.112

Hong JW, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* **321:** 1314. doi:10.1126/science.1160631

Howe FS, Fischl H, Murray SC, Mellor J. 2017. Is H3K4me3 instructive for transcription activation? *Bioessays* **39:** 1–12. doi:10.1002/bies.201670013

Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EE. 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148:** 473–486. doi:10.1016/j.cell.2012.01.030

Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. 2010. Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci* **107:** 2926–2931. doi:10.1073/pnas.0909344107

Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, et al. 2010. Variation in transcription factor binding among humans. *Science* **328:** 232–235. doi:10.1126/science.1183621

Khoueiry P, Girardot C, Ciglar L, Peng PC, Gustafson EH, Sinha S, Furlong EE. 2017. Uncoupling evolutionary changes in DNA sequence, transcription factor occupancy and enhancer activity. *eLife* **6:** e28440. doi:10.7554/eLife.28440

Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, Migliavacca E, Wiederkehr M, Gutierrez-Arcelus M, Panousis NI, et al. 2013. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342:** 744–747. doi:10.1126/science.1242463

Kim JM, Jung YS, Sungur EA, Han KH, Park C, Sohn I. 2008. A copula method for modeling directional dependence of genes. *BMC Bioinformatics* **9:** 225. doi:10.1186/1471-2105-9-225

Kircher M, Xiong C, Martin B, Schubach M, Inoue F, Bell RJA, Costello JF, Shendure J, Ahituv N. 2019. Saturation mutagenesis of 20 disease-associated regulatory elements at single base-pair resolution. *Nat Commun* **10:** 3583. doi:10.1038/s41467-019-11526-w

Knowles DA, Davis JR, Edgington H, Raj A, Favé MJ, Zhu X, Potash JB, Weissman MM, Shi J, Levinson DF, et al. 2017. Allele-specific expression

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

reveals interactions between genetic variation and environment. *Nat Methods* **14:** 699–702. doi:10.1038/nmeth.4298
Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. 2014. Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*. *Nature* **512:** 91–95. doi:10.1038/nature13395
Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. 2005. Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics* **171:** 1813–1822. doi:10.1534/genetics.105.047449
Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S, et al. 2014. Immunogenetics. Chromatin state dynamics during blood formation. *Science* **345:** 943–949. doi:10.1126/science.1256271
Lasserre J, Chung HR, Vingron M. 2013. Finding associations among histone modifications using sparse partial correlation networks. *PLoS Comput Biol* **9:** e1003168. doi:10.1371/journal.pcbi.1003168
Lee N, Kim JM. 2019. Copula directional dependence for inference and statistical analysis of whole-brain connectivity from fMRI data. *Brain Behav* **9:** e01191. doi:10.1002/brb3.1191
Lemos B, Araripe LO, Fontanillas P, Hartl DL. 2008. Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. *Proc Natl Acad Sci* **105:** 14471–14476. doi:10.1073/pnas.0805160105
Levo M, Zalckvar E, Sharon E, Dantas Machado AC, Kalma Y, Lotam-Pompan M, Weinberger A, Yakhini Z, Rohs R, Segal E. 2015. Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res* **25:** 1018–1029. doi:10.1101/gr.185033.114
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26:** 589–595. doi:10.1093/bioinformatics/btp698
Liu X, Li YI, Pritchard JK. 2019. *Trans* effects on gene expression can drive omnigenic inheritance. *Cell* **177:** 1022–1034.e6. doi:10.1016/j.cell.2019.04.014
Long HK, Prescott SL, Wysocka J. 2016. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167:** 1170–1187. doi:10.1016/j.cell.2016.09.018
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8
Lu R, Rogan PK. 2018. Transcription factor binding site clusters identify target genes with similar tissue-wide expression and buffer against mutations. *F1000Res* **7:** 1933. doi:10.12688/f1000research.17363.1
Lucchesi JC, Kuroda MI. 2015. Dosage compensation in *Drosophila*. *Cold Spring Harb Perspect Biol* **7:** a019398. doi:10.1101/cshperspect.a019398
Lynch M, Walsh B. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, MA.
Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* genetic reference panel. *Nature* **482:** 173–178. doi:10.1038/nature10811
Margaritis T, Oreal V, Brabers N, Maestroni L, Vitaliano-Prunier A, Benschop JJ, van Hooff S, van Leenen D, Dargemont C, Géli V, et al. 2012. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3′-end antisense transcription. *PLoS Genet* **8:** e1002952. doi:10.1371/journal.pgen.1002952
Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ. 2014. The roles of *cis*- and *trans*-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res* **24:** 84–95. doi:10.1101/gr.156414.113
Mi H, Vandergriff J, Campbell M, Narechania A, Majoros W, Lewis S, Thomas PD, Ashburner M. 2003. Assessment of genome-wide protein function classification for *Drosophila melanogaster*. *Genome Res* **13:** 2118–2128. doi:10.1101/gr.771603
Mikhaylichenko O, Bondarenko V, Harnett D, Schor IE, Males M, Viales RR, Furlong EEM. 2018. The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev* **32:** 42–57. doi:10.1101/gad.308619.117
Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147:** 1132–1145. doi:10.1016/j.cell.2011.10.023
Moyerbrailean GA, Richards AL, Kurtz D, Kalita CA, Davis GO, Harvey CT, Alazizi A, Watza D, Sorokin Y, Hauff N, et al. 2016. High-throughput allele-specific expression across 250 environmental conditions. *Genome Res* **26:** 1627–1638. doi:10.1101/gr.209759.116
Oktaba K, Zhang W, Lotz TS, Jun DJ, Lemke SB, Ng SP, Esposito E, Levine M, Hilgers V. 2015. ELAV links paused Pol II to alternative polyadenylation in the *Drosophila* nervous system. *Mol Cell* **57:** 341–348. doi:10.1016/j.molcel.2014.11.024
Opgen-Rhein R, Strimmer K. 2007. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst Biol* **1:** 37. doi:10.1186/1752-0509-1-37
Pai AA, Cain CE, Mizrahi-Man O, De Leon S, Lewellen N, Veyrieras JB, Degner JF, Gaffney DJ, Pickrell JK, Stephens M, et al. 2012. The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet* **8:** e1003000. doi:10.1371/journal.pgen.1003000
Pai AA, Pritchard JK, Gilad Y. 2015. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet* **11:** e1004857. doi:10.1371/journal.pgen.1004857
Pal K, Forcato M, Jost D, Sexton T, Vaillant C, Salviato E, Mazza EMC, Lugli E, Cavalli G, Ferrari F. 2019. Global chromatin conformation differences in the *Drosophila* dosage compensated Chromosome X. *Nat Commun* **10:** 5355. doi:10.1038/s41467-019-13350-8
Pradeepa MM, Grimes GR, Kumar Y, Olley G, Taylor GC, Schneider R, Bickmore WA. 2016. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet* **48:** 681–686. doi:10.1038/ng.3550
Reddington JP, Garfield DA, Sigalova OM, Karabacak Calviello A, Marco-Ferreres R, Girardot C, Viales RR, Degner JF, Ohler U, Furlong EE. 2020. Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. *Dev Cell* **55:** 648–664.e9. doi:10.1016/j.devcel.2020.10.009
Reddy TE, Gertz J, Pauli F, Kucera KS, Varley KE, Newberry KM, Marinov GK, Mortazavi A, Williams BA, Song L, et al. 2012. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res* **22:** 860–869. doi:10.1101/gr.131201.111
Robinson MD, McCarthy DJ, Smyth GK. 2010. Edger: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26:** 139–140. doi:10.1093/bioinformatics/btp616
Scholes C, Biette KM, Harden TT, DePace AH. 2019. Signal integration by shadow enhancers and enhancer duplications varies across the *Drosophila* embryo. *Cell Rep* **26:** 2407–2418.e5. doi:10.1016/j.celrep.2019.01.115
Schor IE, Degner JF, Harnett D, Cannavò E, Casale FP, Shim H, Garfield DA, Birney E, Stephens M, Stegle O, et al. 2017. Promoter shape varies across populations and affects promoter evolution and expression noise. *Nat Genet* **49:** 550–558. doi:10.1038/ng.3791
Spitz F, Furlong EE. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13:** 613–626. doi:10.1038/nrg3207
Spivakov M, Akhtar J, Kheradpour P, Beal K, Girardot C, Koscielny G, Herrero J, Kellis M, Furlong EE, Birney E. 2012. Analysis of variation at transcription factor binding sites in *Drosophila* and humans. *Genome Biol* **13:** R49. doi:10.1186/gb-2012-13-9-r49
Starz-Gaiano M, Cho NK, Forbes A, Lehmann R. 2001. Spatially restricted activity of a *Drosophila* lipid phosphatase guides migrating germ cells. *Development* **128:** 983–991.
Sun W, Gao Q, Schaefke B, Hu Y, Chen W. 2018. Pervasive allele-specific regulation on RNA decay in hybrid mice. *Life Sci Alliance* **1:** e201800052. doi:10.26508/lsa.201800052
Tirosh I, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324:** 659–662. doi:10.1126/science.1169766
Turner LM, Chuong EB, Hoekstra HE. 2008. Comparative analysis of testis protein evolution in rodents. *Genetics* **179:** 2075–2089. doi:10.1534/genetics.107.085902
Uhl JD, Zandvakili A, Gebelein B. 2016. A Hox transcription factor collective binds a highly conserved distal-less *cis*-regulatory module to generate robust transcriptional outcomes. *PLoS Genet* **12:** e1005981. doi:10.1371/journal.pgen.1005981
Urban J, Kuzu G, Bowman S, Scruggs B, Henriques T, Kingston R, Adelman K, Tolstorukov M, Larschan E. 2017. Enhanced chromatin accessibility of the dosage compensated *Drosophila* male X-chromosome requires the CLAMP zinc finger protein. *PLoS One* **12:** e0186855. doi:10.1371/journal.pone.0186855
Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM, Orioli A, Wiederkehr M, Panousis NI, Yurovsky A, et al. 2015. Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162:** 1039–1050. doi:10.1016/j.cell.2015.08.001
Waymack R, Fletcher A, Enciso G, Wunderlich Z. 2020. Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *eLife* **9:** e59351. doi:10.7554/eLife.59351
Wilczyński B, Furlong EE. 2010. Dynamic CRM occupancy reflects a temporal map of developmental progression. *Mol Syst Biol* **6:** 383. doi:10.1038/msb.2010.35

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430:** 85–88. doi:10.1038/nature02698

Wong ES, Schmitt BM, Kazachenka A, Thybert D, Redmond A, Connor F, Rayner TF, Feig C, Ferguson-Smith AC, Marioni JC, et al. 2017. Interplay of *cis* and *trans* mechanisms driving transcription factor binding and gene expression evolution. *Nat Commun* **8:** 1092. doi:10.1038/s41467-017-01037-x

Xiong N, Kang C, Raulet DH. 2002. Redundant and unique roles of two enhancer elements in the TCRγ locus in gene regulation and γδ T cell development. *Immunity* **16:** 453–463. doi:10.1016/S1074-7613(02)00285-6

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2015. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518:** 556–559. doi:10.1038/nature13994

Zheng W, Zhao H, Mancera E, Steinmetz LM, Snyder M. 2010. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464:** 1187–1191. doi:10.1038/nature08934

Zhu J, Sanborn JZ, Diekhans M, Lowe CB, Pringle TH, Haussler D. 2007. Comparative genomics search for losses of long-established genes on the human lineage. *PLoS Comput Biol* **3:** e247. doi:10.1371/journal.pcbi.0030247

Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EE. 2009. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462:** 65–70. doi:10.1038/nature08531

Received May 21, 2020; accepted in revised form December 9, 2020.

Downloaded from genome.cshlp.org on September 22, 2021 - Published by Cold Spring Harbor Laboratory Press

Cis-acting variation is common across regulatory layers but is often buffered during embryonic development

Swann Floc'hlay, Emily S. Wong, Bingqing Zhao, et al.

Genome Res. 2021 31: 211-224 originally published online December 11, 2020
Access the most recent version at doi:10.1101/gr.266338.120

Supplemental Material http://genome.cshlp.org/content/suppl/2021/01/19/gr.266338.120.DC1

References This article cites 86 articles, 31 of which can be accessed free at:
http://genome.cshlp.org/content/31/2/211.full.html#ref-list-1

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at http://creativecommons.org/licenses/by/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

© 2021 Floc'hlay et al.; Published by Cold Spring Harbor Laboratory Press