Molecular Evolution across Mouse Spermatogenesis

EMILY E. K. KOPANIA[1], ERICA L. LARSON[2], COLIN CALLAHAN[1], SARA KEEBLE[1], JEFFREY M. GOOD[1]

Abstract

Genes involved in spermatogenesis tend to evolve rapidly, but we still lack a clear understanding of how different components of molecular evolution vary across this complex developmental process. We used fluorescence activated cell sorting (FACS) to generate expression data for both early (meiotic) and late (postmeiotic) cell types across thirteen inbred strains of mice (*Mus*) spanning ~7 million years of evolution. We used these comparative developmental data to investigate the evolution of lineage-specific expression, protein-coding sequences, and expression levels. We found increased lineage specificity and more rapid protein-coding and expression divergence during late spermatogenesis, suggesting that signatures of rapid testis molecular evolution are punctuated across sperm development. Despite strong overall developmental parallels in these components of molecular evolution, protein and expression divergences were only weakly correlated across genes. We detected more rapid protein evolution on the X chromosome relative to the autosomes, while X-linked gene expression tended to be relatively more conserved likely reflecting chromosome-wide regulatory constraints. Using allele-specific FACS expression data from crosses between four strains, we found that the relative contributions of different regulatory mechanisms also differed between cell-types. Genes showing *cis*-regulatory changes were more common late in spermatogenesis, and tended to be associated with larger differences in expression levels and greater expression divergence between species. In contrast, genes with *trans*-acting changes were more common early and tended to be more conserved across species. Our findings advance understanding of gene evolution across spermatogenesis and underscore the fundamental importance of developmental context in molecular evolutionary studies.

Introduction

Mature sperm are the most morphologically diverse animal cell type, likely as a consequence of intense selection on sperm form and function (Pitnick, et al. 2009; Pizzari and Parker 2009). Genes involved in spermatogenesis also tend evolve rapidly (Wyckoff, et al. 2000; Torgerson, et al. 2002; Swanson, et al. 2003; Good and Nachman 2005; Turner, et al. 2008; Larson, et al. 2016; Ahmed-Braimah, et al. 2017; Whittington, et al. 2017; Finseth and Harrison 2018), suggesting that pervasive sexual selection also shapes molecular evolution (Swanson and Vacquier 2002; Harrison, et al. 2015). However, direct genotype-to-phenotype connections remain elusive for primary sexually selected traits, and there are additional evolutionary forces acting during spermatogenesis that shape overall patterns of molecular evolution (Good and Nachman 2005; Burgoyne, et al. 2009; Dean, et al. 2009; Meisel 2011; Walters and Harrison 2011; Schumacher, et al. 2014; Larson, et al. 2016; Schumacher and Herlyn 2018). For example, many spermatogenesis genes are highly specialized (Eddy 2002; Chalmel, et al. 2007; Green, et al. 2018), which can relax pleiotropic constraint and contribute to rapid evolution even in the absence of positive directional selection (Winter, et al. 2004; Larracuente, et al. 2008; Meisel 2011; Walters and Harrison 2011). Other components of spermatogenesis are highly conserved because small disruptions can lead to infertility (Burgoyne, et al. 2009). Thus, spermatogenesis genes are likely to experience strong and sometimes contradictory evolutionary pressures. Understanding how these processes interact to shape molecular evolution across spermatogenesis is essential to understanding how natural selection shapes the genetic determinants of male fertility.

There are many components or levels of molecular evolution, spanning from protein sequence changes to differences in gene expression level, timing, and developmental specificity (King and Wilson 1975; Betrán, et al. 2002; Wray, et al. 2003; Larracuente, et al. 2008; Kaessmann 2010; Piasecka, et al. 2013; Cridland, et

[1] DIVISION OF BIOLOGICAL SCIENCES, UNIVERSITY OF MONTANA, MISSOULA, MT 59812, USA
[2] DEPARTMENT OF BIOLOGICAL SCIENCES, UNIVERSITY OF DENVER, DENVER, CO 80208, USA
CORRESPONDING AUTHOR(S): JEFFREY M. GOOD JEFFREY.GOOD@UMONTANA.EDU; EMILY E. K. KOPANIA EMILY.KOPANIA@UMCONNECT.UMT.EDU

al. 2020). Many of these components have been shown to evolve relatively rapidly during spermatogenesis (Meiklejohn, et al. 2003; Ranz, et al. 2003; Khaitovich, et al. 2005; Voolstra, et al. 2007; Brawand, et al. 2011; Kousathanas, et al. 2014; Harrison, et al. 2015; Vicens, et al. 2017; Cridland, et al. 2020; Sánchez-Ramírez, et al. 2021), and generally trend towards increased divergence during the later stages of development (Good and Nachman 2005; Piasecka, et al. 2013; Larson, et al. 2016). Novel genes disproportionately arise with testis-specific expression (Betrán, et al. 2002; Levine, et al. 2006; Kaessmann 2010; Assis and Bachtrog 2013; Zhao, et al. 2014; Schroeder, et al. 2019; Cridland, et al. 2020; Lange, et al. 2021), likely as a consequence of the more permissive regulatory environment of the later stages of sperm development (Kaessmann 2010; Soumillon, et al. 2013). Likewise, the later stages of spermatogenesis tend to be enriched for novel testis-specific genes (Eddy 2002; Chalmel, et al. 2007; Green, et al. 2018). These developmental signatures of novelty and specialization are further reflected in patterns of increased divergence of protein sequences (Good and Nachman 2005; Kousathanas, et al. 2014) and expression levels (Larson, et al. 2016) between species during the later stages of sperm development. Parallel signatures of rapid molecular evolution likely reflect both relaxed constraints during the late stages of spermatogenesis, and enhanced positive selection on late-developing sperm phenotypes (Eddy 2002; Good and Nachman 2005; Chalmel, et al. 2007; Larracuente, et al. 2008; Abzhanov 2013; Piasecka, et al. 2013; Larson, et al. 2016; Cutter and Bundus 2020). However, it remains unclear how strongly different forms of molecular evolution are correlated. For example, changes in gene expression may often be cell or stage-specific and therefore may be less pleiotropic than protein-coding changes (Wray, et al. 2003; Wray 2007; Carroll 2008; Piasecka, et al. 2013; Cutter and Bundus 2020). This pleiotropic constraint hypothesis primarily applies to *cis*-regulatory changes, which likely affect one gene, whereas *trans*-regulatory changes can affect many genes across multiple cell types (Wray, et al. 2003; Carroll 2008; Cutter and Bundus 2020).

The X chromosome provides a compelling example of how the conflicting selective pressures acting on spermatogenesis may shape different components of molecular evolution. Theory predicts that the X chromosome should evolve more rapidly than the autosomes due to a lower effective population size (i.e., increased genetic drift) and more efficient selection resulting from hemizygosity in males, which immediately exposes recessive beneficial and deleterious mutations to selection (Charlesworth, et al. 1987; Vicoso and Charlesworth 2009). Consistent with these predictions, protein-coding evolution tends to be faster on the X chromosome compared to the autosomes in several taxa, and this effect is often strongest for genes with male-biased expression (Charlesworth, et al. 1987; Khaitovich, et al. 2005; Baines and Harr 2007; Baines, et al. 2008; Meisel and Connallon 2013; Parsch and Ellegren 2013; Larson, et al. 2016). Novel genes tend to arise more often on the X chromosome, and these are often expressed during spermatogenesis (Betrán, et al. 2002; Levine, et al. 2006; Kaessmann 2010). There is also some evidence for rapid expression evolution on the X chromosome in flies and mammals (Khaitovich, et al. 2005; Brawand, et al. 2011; Meisel, et al. 2012; Coolon, et al. 2015), but X-linked expression in mice appears conserved relative to autosomal genes expressed during the later stages of spermatogenesis (Larson, et al. 2016). Stage-specific differences in relative rates of expression evolution on the X chromosome may result from the unique regulatory pattern that the sex chromosomes undergo during mammalian spermatogenesis. In males, the X chromosome is inactivated early in meiosis (i.e., meiotic sex chromosome inactivation, MSCI; McKee and Handel 1993; Turner 2015) and remains partially repressed during the postmeiotic haploid stages of sperm development (i.e., postmeiotic sex chromosome repression, PSCR; Namekawa, et al. 2006; Sin, et al. 2012; Larson, et al. 2016).

These stage-specific patterns highlight the importance of studying specific components of molecular evolution in a developmental framework (Figure 1*A*; Larson, et al. 2018a; Cutter and Bundus 2020). However, studies of molecular evolution have primarily focused on pairwise contrasts across nuanced aspects of tissue development (Good and Nachman 2005; Larson, et al. 2016), or examined protein-coding versus regulatory evolution in whole tissues (Khaitovich, et al. 2005; Voolstra, et al. 2007; Larson, et al. 2016; Mack, et al. 2016; Vicens, et al. 2017; Cridland, et al. 2020), without combining both in a phylogenetic framework. Relying on whole tissue expression comparisons may be particularly problematic for spermatogenesis, because differences in testis composition are expected to evolve rapidly between species (Ramm and Schärer 2014; Yapar, et al. 2021) and may confound patterns of expression level divergence (Good, et al. 2010; Larson, et al. 2016; Hunnicutt, et al. 2021). Nonetheless, collection of stage or cell-specific expression data remains technically demanding (Getun, et al. 2011; da Cruz, et al. 2016; Green, et al. 2018; Cao, et al. 2020), likely limiting widespread use in comparative studies. As a consequence, most evolutionary studies of gene expression have relied on whole tissue comparisons between closely-related species pairs, instead of using more powerful phylogenetic approaches (Rohlfs and Nielsen 2015; Dunn, et al. 2018).

In this study, we use a comparative developmental approach to gain a more comprehensive understanding of molecular evolution across spermatogenesis in house mice (*Mus*). Mice are the predominant laboratory model for mammalian reproduction (Phifer-Rixey and Nachman 2015; Firman 2020), with abundant genomic resources (Keane, et al. 2011; Thybert, et al. 2018), and established wild-derived inbred strains that can be crossed to resolve mechanisms underlying expression divergence (i.e., *cis*- versus *trans*-regulatory changes; Mack, et al. 2016). Mice also show divergence in sperm head morphologies across closely related species (Albrechtová, et al. 2014; Skinner, et al. 2019) and experience sperm competition in the wild (Dean, et al. 2006), providing a compelling system for understanding the evolution of spermatogenesis.
We used fluorescence activated cell sorting (FACS) to

resolve patterns of gene expression in two enriched spermatogenic cell populations across several mouse strains, species, and cross-types (Figure 1*A*). Our study used two main comparisons. First, we evaluated divergence in spermatogenic protein sequences and gene expression levels across thirteen inbred strains of mice, including two subspecies of the house mouse (*Mus musculus*) and two other *Mus* species spanning seven million years of evolution (Figure 1*B*; Chevret, et al. 2005). Second, we used published data from reciprocal crosses between a subset of these inbred strains to employ allele-specific expression analyses to resolve the relative contribution of *cis*- versus *trans*-regulatory changes to expression divergence.

We used these data to address five main questions: (i) Is gene expression more lineage-specific during late spermatogenesis? (ii) Do protein-coding sequences and gene expression levels evolve faster during the later stages of spermatogenesis? (iii) Is the rate of molecular evolution elevated on the X chromosome compared to the autosomes, and does this relationship change across spermatogenesis? (iv) To what extent are protein-coding and gene expression divergence correlated, and does this relationship change across developmental stages? (v) Are there differences in the relative contributions of regulatory mechanisms (*cis*- versus *trans*-regulatory changes) across spermatogenesis?

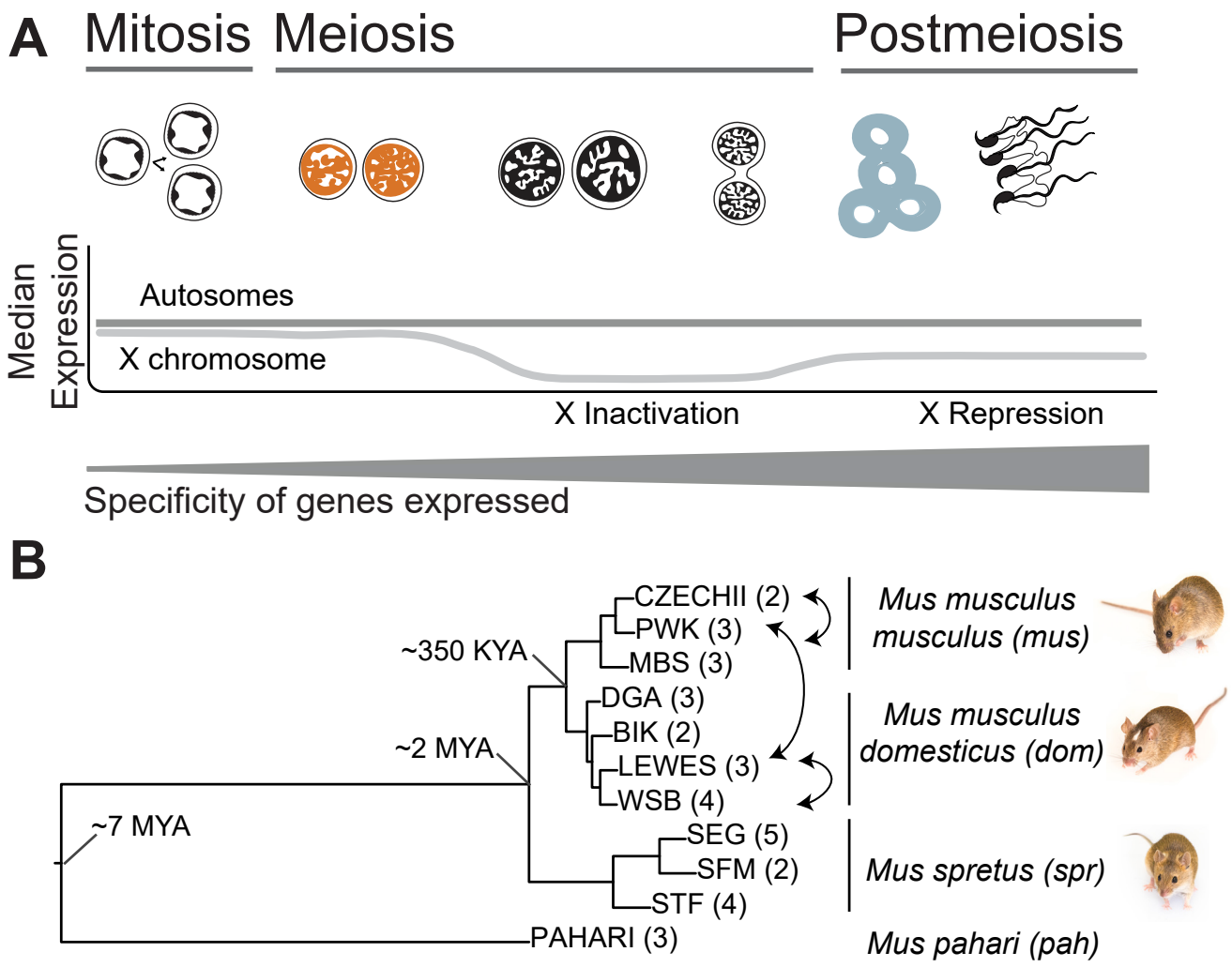

Figure 1. (*A*) Predictive framework depicting the major stages of spermatogenesis and expected relative expression levels of the X chromosome and autosomes at each stage (Namekawa, et al. 2006). The two cell populations used in this study are leptotene-zygotene ("early", second from left, orange) and round spermatids ("late", second from right, blue). The relative thickness of the gray bar represents the predicted cell type specificity at each stage (Eddy 2002; Chalmel, et al. 2007; Larson, et al. 2016; Green, et al. 2018). (*B*) Maximum likelihood tree of concatenated exome data from the four *Mus* species or subspecies used in this study: *Mus musculus musculus* (*mus*), *Mus musculus domesticus* (*dom*), *Mus spretus* (*spr*), *Mus pahari* (*pah*). Tips are labeled with the inbred strains from each lineage, with select crosses used to generate F1 hybrids indicated with arrows. Number of individuals sampled for each strain indicated in parentheses. Approximate divergence times are placed at each major node (Chevret, et al. 2005). All nodes had 100% bootstrap support.

Results

Spermatogenesis Gene Expression by Cell Type and Lineage

We collected spermatogenesis expression data from 34 mice representing four different species or subspecies: *Mus musculus musculus*, *Mus musculus domesticus*, *Mus spretus*, and *Mus pahari*. We will use the abbreviations *mus*, *dom*, *spr*, and *pah* to reference the four major groups, and refer to all taxa as "lineages" for concision (Figure 1*B*). For each sample, we generated expression data for two spermatogenic cell types, an early meiotic cell type (leptotene-zygotene cells from early prophase of meiosis I, hereafter "early") and a post-meiotic cell type (round spermatids, hereafter "late"). We identified 23,164 one-to-one orthologs, including both protein-coding and non-protein-coding genes, that were annotated in all four mouse lineages and the mouse reference (GRCm38), and only included these genes in our analyses. From this set of orthologs, we identified sets of expressed genes as those with an FPKM > 1 in all samples of a given cell type. Expression variance cleanly separated samples by cell type and lineage (supplementary figure S1), indicating successful enrichment of different cell types. Most expressed genes were detected in both the early and late cell types (Table 1). However, approximately one third of the detected genes were preferentially expressed or "induced" in a given cell type (transcripts with > 2X median expression level in one cell type compared to the other across all lineages; Table 1). We also identified the subset of genes expressed in our dataset that show testis-specific expression based on published multi-tissue expression data (Chalmel, et al. 2007). We found that 493 testis-specific genes were induced late, while only 65 testis-specific genes were induced early (Table 1), consistent with previous studies showing increased specificity late in spermatogenesis (Eddy 2002; Larson, et al. 2016; Green, et al. 2018). To further distinguish experimental noise from biologically meaningful expression, we also used a Bayesian approach to determine if a gene was "active" in a tissue or cell type (Thompson, et al. 2020) and found broad overlap with genes in the expressed dataset (Table 1). Using the same framework, we identified genes showing evidence for lineage-specific expression ("active" in a single

Table 1. Counts of genes in each dataset and spermatogenic cell type. Numbers in parentheses represent the percent of genes in the "active" datasets that were also in the "expressed" dataset. TS = testis-specific (inferred from Chalmel, et al. 2007)

	early (spermatocytes)	late (round spermatids)	both early and late
expressed	9570	8986	7670
induced	3375	2769	0
testis-specific (TS)	544	655	524
induced and TS	65	493	0
active (dom)	8206 (98.2%)	8581 (90.4%)	6355
active (mus)	8782 (97.5%)	10098 (83.4%)	7289
active (spr)	8728 (97.1%)	9509 (86.0%)	7227
active (pah)	8124 (97.6%)	9563 (83.9%)	6682

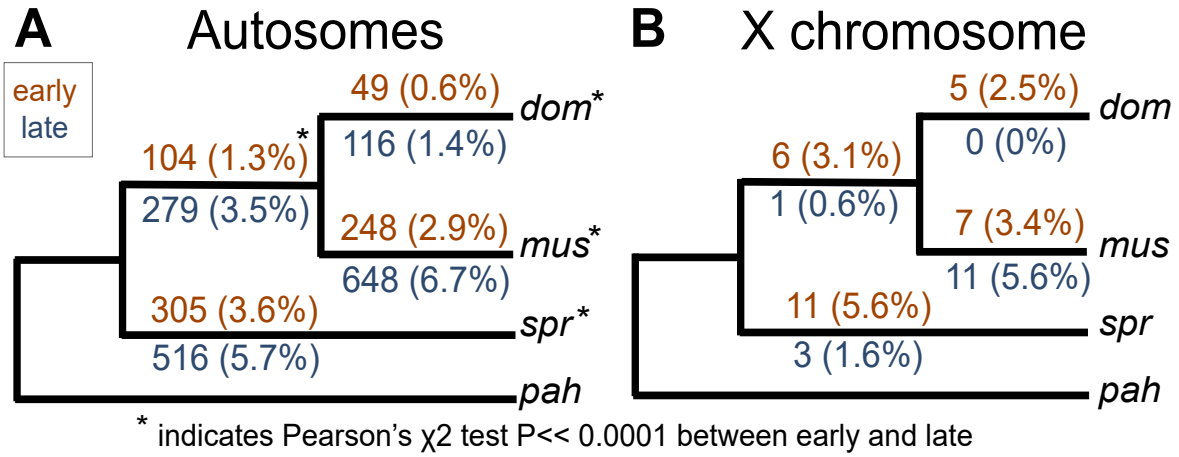

Figure 2. Number of genes that were lineage-specific on each internal branch of the mouse phylogeny used in this study. Numbers in parentheses are the percent of active genes that were lineage-specific. Results are presented separately for the autosomes (*A*) and X chromosome (*B*). Orange values above each branch represent the early cell type and blue values below represent the late cell type. Asterisks indicate a significant difference between early and late on that branch based on a Pearson's χ2 test.

lineage or subset of lineages). These broad categories of cell-type induced and testis-specific genes provide general metrics for the degree of gene specialization across stages of spermatogenesis.

We found that rates of lineage-specificity were low overall, but relatively more common for autosomal genes active during late spermatogenesis (Pearson's $\chi 2$ test; *dom*: P << 0.0001, *mus*: P << 0.0001, *spr*: P << 0.0001, *dom-mus* common ancestor: P << 0.0001; Figure 2*A*). X-linked genes showed no significant differences in lineage-specificity between early and late cell types (Figure 2*B*), which could reflect a lack of specialization on the sex chromosomes, or just reduced power to detect differences between cell types given small sample sizes. Relatively few genes were lineage-specific in both cell types, and all of these genes were autosomal (*dom*: 9 genes, *mus*: 24 genes, *spr*: 24 genes, *dom-mus*: 21 genes). Similar results were obtained when using a log fold-change (logFC) approach to identify lineage-specific genes (supplementary figure S2). Collectively, these results suggest that lineage-specific expression of spermatogenic genes is relatively rare overall at these shallow phylogenetic scales, but more likely to arise later in spermatogenesis.

Greater Protein-Coding and Gene Expression Divergence during Late Spermatogenesis

Having detected evidence for subtle increases in lineage-specificity late in spermatogenesis, we next tested if rates of protein sequence evolution (dN/dS) and expression level divergence were also elevated during the postmeiotic stage, as has been reported previously in house mice using less extensive expression data (Good and Nachman 2005; Larson, et al. 2016). For these comparisons, we first evaluated the autosomes and X chromosome separately. Genes induced late in spermatogenesis showed significantly higher rates of protein-coding divergence on both the autosomes (n = 2046 genes induced early, median dN/dS = 0.11; n = 1711 genes induced late, median dN/dS = 0.20; Wilcoxon rank sum test P << 0.0001) and the X chromosome (n = 54 genes induced early, median dN/dS = 0.25; n = 61 genes induced late, median dN/dS = 0.41; Wilcoxon rank sum test P = 0.049; Figure 3*A*, supplementary tables S1 and S2). Testis-specific genes showed elevated dN/dS overall, but there was no significant difference between genes expressed early and late for the autosomes (n = 350 genes expressed early, median dN/dS = 0.28; n = 424 genes expressed late, median dN/dS = 0.30; Wilcoxon rank sum test P = 1) or the X chromosome (n = 16 genes expressed early; median dN/dS = 0.59; n = 24 genes expressed late, median dN/dS = 0.58; Wilcoxon rank sum test P = 1). We were able to calculate a dN/dS value for 489 testis-specific genes expressed in our dataset, and most were expressed in both cell types. However, most of these testis-specific genes were preferentially expressed in the late cell type, with 348 genes induced late, representing about 20% of all genes induced late for which we were able to calculate a dN/dS value. Taken together, these results confirm that tissue specificity plays an important role in the rapid protein-coding divergence of spermatogenic genes, and that most of this signature involves genes induced in postmeiotic spermatogenesis.

We used the Expression Variance and Evolution (EVE) model (Rohlfs and Nielsen 2015) to quantify expression level divergence, which uses a phylogenetic ANOVA to estimate divergence while controlling for phylogenetic relatedness and variance within lineages (see Methods for details). We report expression divergence from EVE as -log($beta_i$), where $beta_i$ is a metric from EVE that represents the ratio of within-lineage variance to between-lineage evolutionary divergence. By taking the negative log, higher positive numbers correspond to greater divergence between lineages.

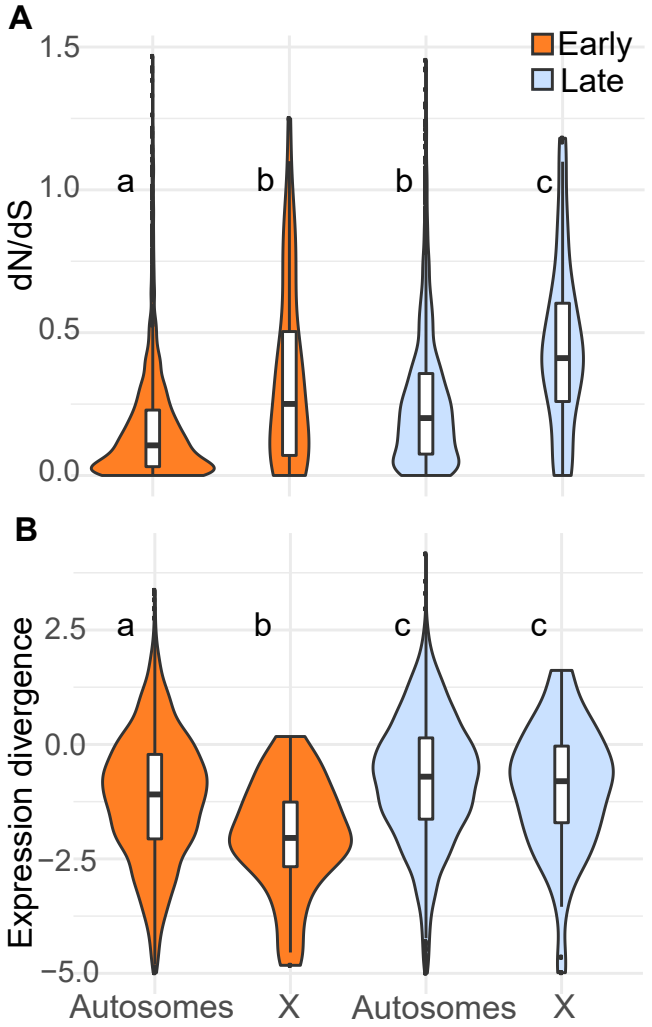

Figure 3. (*A*) Protein-coding and (*B*) expression divergence on the autosomes and X chromosome for genes induced in each cell type. Expression divergence values on the y-axis are -log($beta_i$), where $beta_i$ is the measure of expression divergence from EVE. Higher values on the y-axis represent higher divergence. The center of each violin plot is a standard boxplot, with the center horizontal line representing the median divergence value. The violins show the probability density of divergence values for each group. A wider part of the violin at a given value means genes expressed in that group are more likely to have that divergence value. The letters above each violin indicate significant differences between the cell types and chromosome types based on a Wilcoxon rank-sum test.

Expression divergence was higher for genes induced late in spermatogenesis on both the autosomes (n = 2461 genes induced early, median EVE divergence = -1.09; n = 2305 genes induced late, median EVE divergence = -0.70; Wilcoxon rank sum test P << 0.0001) and X chromosome (n = 44 genes induced early, median EVE divergence = -2.04; n = 68 genes induced late, median EVE divergence = -0.80; Wilcoxon rank sum test P = 0.00019; Figure 3*B*). This pattern held when looking at all expressed genes and testis-specific genes (supplementary table S3). We also found higher divergence late for expressed and induced autosomal genes (supplementary table S3) based on pairwise expression divergences using logFC and the metric from Meisel, et al. (2012). The pairwise framework did not give a consistent pattern on the X chromosome. When looking at all genes, most pairwise comparisons showed higher divergence late, but induced genes showed no difference between early and late spermatogenesis for most comparisons. However, the *dom* versus *spr* comparison had lower divergence late for all expressed genes and induced genes (supplementary table S3).

Collectively, we found strong evidence for more rapid protein-coding and gene expression level divergence during postmeiotic spermatogenesis, suggesting that these general patterns hold after controlling for phylogeny and at deeper divergence levels than had previously been shown in mice (Good and Nachman 2005; Larson, et al. 2016). Note that despite our expanded phylogenetic sample, we still lacked the power to determine if more rapid molecular divergence is due to positive directional selection (supplementary figure S3).

Weak Positive Correlation between Gene Expression and Protein-Coding Divergence

Given strong developmental parallels for protein-coding evolution and expression divergence, we next tested for more general relationships between protein-coding and expression divergence across sets of genes expressed or induced during spermatogenesis (supplementary figure S4, supplementary table S4). Across all autosomal genes expressed early, there was a weak positive correlation between dN/dS and pairwise expression divergence (ρ = 0.13-0.17, Spearman's rank correlation P << 0.0001). For induced genes, this correlation was weaker but still significant (ρ = 0.07-0.11, Spearman's rank correlation P < 0.05). For the late cell type, there was also a weak positive correlation between pairwise expression divergence and dN/dS on the autosomes, but the correlation was weaker than that seen in the early cell type (ρ = 0.03-0.05, Spearman's rank correlation P < 0.05). There was no significant correlation for the set of genes induced late. When looking only at genes with evidence for positive directional selection at the protein-coding level after correction for multiple tests (366 genes), the correlation was stronger on the autosomes late for the *dom* vs *spr* (n = 250 genes, ρ = 0.17, Spearman's rank correlation P = 0.02) and *mus* vs *spr* comparisons (n = 249 genes, ρ = 0.18, Spearman's rank correlation P << 0.0001). When comparing dN/dS to EVE expression divergence, we only saw a significant positive correlation for genes expressed late that were also under positive selection at the protein-coding level (n = 160 genes, ρ = 0.18, Spearman's rank correlation P = 0.04). In summary, we tended to observe a positive relationship between protein-coding and expression level divergence, but the strength of this relationship was weak and varied by gene set and divergence metric.

Faster-X Protein-Coding but Not Gene Expression Evolution

In addition to comparisons between spermatogenesis cell types, we compared relative rates of molecular evolution between X-linked and autosomal genes within a cell type. We found that protein-coding divergence was higher on the X chromosome, both early and late, across all gene sets (Figure 3*A*, supplementary tables S1 and S2) consistent with several previous studies (Khaitovich, et al. 2005; Baines, et al. 2008; Meisel and Connallon 2013; Kousathanas, et al. 2014; Larson, et al. 2016). For expression evolution, we found lower divergence on the X chromosome early using EVE (n = 2461 autosomal genes, median EVE divergence = -1.09; n = 44 X-linked genes, median EVE divergence = -2.04; Wilcoxon rank sum test P =0.00015; Figure 3*B*), but higher X-linked divergence when using pairwise comparisons (supplementary table S3). A major difference between these approaches is that EVE calculates divergence across a phylogeny, so genes that show divergent expression levels in one lineage may still be conserved across the entire phylogeny. We detected significant correlations between pairwise divergence values for different pairwise comparisons on the autosomes, and during late spermatogenesis, but lower or non-significant correlations on the X early (Table 2). Thus, many genes on the X chromosome expressed early show relatively high divergence between two particular lineages, but lower divergence across other pairwise comparisons and across the phylogeny as a whole. This lineage-specific variance underscores the importance of evaluating gene expression divergence in a phylogenetic framework (Rohlfs and Nielsen 2015; Dunn, et al. 2018).

In late spermatogenic cells (round spermatids), X-linked expression divergence was similar to or lower than on the autosomes depending on the contrast and approach. With the EVE approach, we detected a similar degree of divergence on the X chromosome compared to the autosomes late (n = 2305 autosomal genes, median EVE divergence = -0.70; n = 68 X-linked genes, median EVE divergence = -0.80; Wilcoxon rank sum test P = 0.34; Figure 3*B*), while pairwise comparisons gave mixed results, depending on which two lineages were compared (supplementary table S3). There were proportionally fewer differentially expressed genes on the X chromosome (Figure 4, supplementary figure S5), and this pattern was strongest for the more closely related comparisons (hypergeometric test; *mus* versus *dom* P << 0.0001, *spr* versus *dom* P << 0.0001, *spr* versus *mus* P << 0.0001). Across all metrics of expression divergence and both developmental stages, there was no evidence for pervasive faster-X gene expression level evolution.

Table 2. Correlation between pairwise expression divergence values for all possible pairwise comparisons. Numbers presented are ρ values from a Spearman's rank correlation test. We tested for correlations in pairwise expression divergence value among induced genes in each stage and chromosome group (early X, early autosomal, late X, and late autosomal). Gray boxes indicate no significant correlation between pairwise divergence values after FDR correction (Spearman's rank correlation P > 0.05). Bolded values indicate the lowest Spearman's ρ value for each pairwise comparison across the four stages and chromosome groups.

		dom vs *mus*	*dom* vs *spr*	*mus* vs *spr*	*dom* vs *pah*	*mus* vs *pah*
Early,	*dom* vs *spr*	0.34				
X-linked	*mus* vs *spr*	**0.07**	**0.28**			
	dom vs *pah*	**0.07**	**0.14**	**0.19**		
	mus vs *pah*	**0.16**	**0.10**	**0.03**	**0.62**	
	spr vs *pah*	**0.14**	0.27	**0.16**	0.58	0.67
Early,	*dom* vs *spr*	**0.32**				
autosomal	*mus* vs *spr*	0.32	0.61			
	dom vs *pah*	0.28	0.28	0.27		
	mus vs *pah*	0.29	0.26	0.30	0.74	
	spr vs *pah*	0.24	0.32	0.34	**0.55**	**0.57**
Late,	*dom* vs *spr*	0.36				
X-linked	*mus* vs *spr*	0.50	0.45			
	dom vs *pah*	0.20	0.23	0.22		
	mus vs *pah*	0.28	0.28	0.36	0.74	
	spr vs *pah*	0.15	**0.20**	0.20	0.73	0.72
Late,	*dom* vs *spr*	0.35				
autosomal	*mus* vs *spr*	0.37	0.59			
	dom vs *pah*	0.30	0.33	0.30		
	mus vs *pah*	0.30	0.30	0.33	0.76	
	spr vs *pah*	0.25	0.32	0.33	0.64	0.63

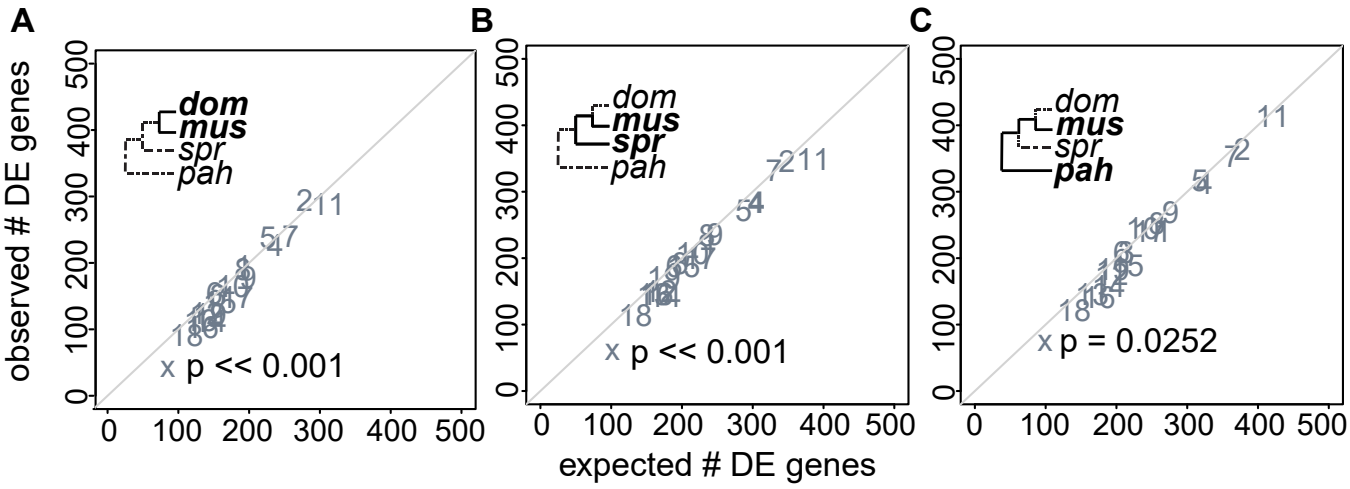

Figure 4. Observed versus expected number of genes differentially expressed (DE) in late spermatogenesis for three pairwise comparisons at different levels of evolutionary divergence: (*A*) *dom* versus *mus*, (*B*) *spr* versus *mus*, and (*C*) *pah* versus *mus*. Each point represents a different chromosome. The diagonal line is the one-to-one line at which the observed number of DE genes equals the expected number. P-values are shown for the X chromosome only. They are based on a hypergeometric test for enrichment and corrected for multiple tests using a false discovery rate correction. A significant p-value indicates that the observed number of DE genes is different from the expected number.

Relative Contributions of *cis*- and *trans*-Regulatory Evolution Vary across Spermatogenesis

Having shown differences in expression divergence between cell types, we next asked if there were differences in the types of regulatory mutations (e.g., *cis*- versus *trans*-regulatory changes) underlying expression divergence in each cell type. We used whole testis and FACS-sorted data from reciprocal crosses between house mouse subspecies (*dom* x *mus*) to estimate allele-specific expression (ASE) and assign genes to eight different regulatory categories: *cis*, *trans*, *cis* X *trans*, compensatory, *cis* + *trans* opposite, *cis* + *trans* same, other, and conserved (Coolon, et al. 2014; Mack, et al. 2016). Across all cell types and genotypes, 50-90% of genes were conserved, or did not have significant evidence for differential expression (DE) between subspecies or ASE between alleles. We excluded these conserved genes when calculating the percent of genes in each regulatory category.

Comparing the two spermatogenic stages, we saw striking differences in the proportions of genes within each regulatory category (Figure 5, supplementary table S5). *Trans* was more common than *cis* early, whereas *trans* and *cis* made up a similar proportion of regulatory changes late (Figure 5, supplementary table S5). Compensatory changes (compensatory and *cis*+*trans* opposite) were more common than reinforcing (*cis*+*trans* same) in both cell types, but there was a higher relative proportion of reinforcing late (Figure 5, supplementary table S5). Figure 5 shows the results for the *dom* (LEWES)♀ X *mus* (PWK)♂ hybrid, which we chose to focus on because these F1 hybrids are more fertile (Good, et al. 2008) and therefore less likely to have misexpressed genes due to hybrid incombatibilities (Good, et al. 2010; Mack, et al. 2016; Larson, et al. 2017). However, the subfertile reciprocal hybrid genotype *mus* (PWK)♀ X *dom* (LEWES)♂ showed similar overall proportions of genes in each regulatory category. The proportions of different regulatory mechanisms in whole testes were more similar to the late cell type (supplementary table S5), consistent with previous studies showing high overlap in expression profiles between whole testes and spermatid stage cells (Soumillon, et al. 2013). We also repeated our analyses using pure strain (LEWES and PWK) expression data from our phylogenetic expression dataset to determine differences in parental strain expression levels, which gave the same overall patterns (supplementary table S5).

We also evaluated the relative contributions of regulatory mechanisms within each *M. musculus* subspecies to test if these proportions changed with evolutionary divergence. To do this, we used expression data from within-subspecies F1s (WSB X LEWES and CZECHII X PWK) and from the respective parental inbred strains. Consistent with the more divergent F1 hybrids, there was more *trans* than *cis* early and about 30% *trans* late. In *dom*, there was a lower proportion of *trans* early compared to the results using F1 hybrids. In both species, there was less contribution from *cis*-regulatory changes late compared to the F1 hybrids (supplementary table S5). There were more compensatory than reinforcing mutations in both cell types, but *dom* showed a higher relative proportion of reinforcing mutations early compared to *mus* and the F1 hybrids (supplementary table S5). In summary, we found differences between early and late spermatogenesis in the types of regulatory mutations contributing to expression divergence, with a proportionally higher contribution of *trans*-regulatory changes early. This pattern was consistent across different degrees of evolutionary divergence and between reciprocal crosses.

***cis*-Regulatory Changes Tended to Have Larger Effects on Expression Level Divergence**

Given that *trans*-regulatory changes were proportionally

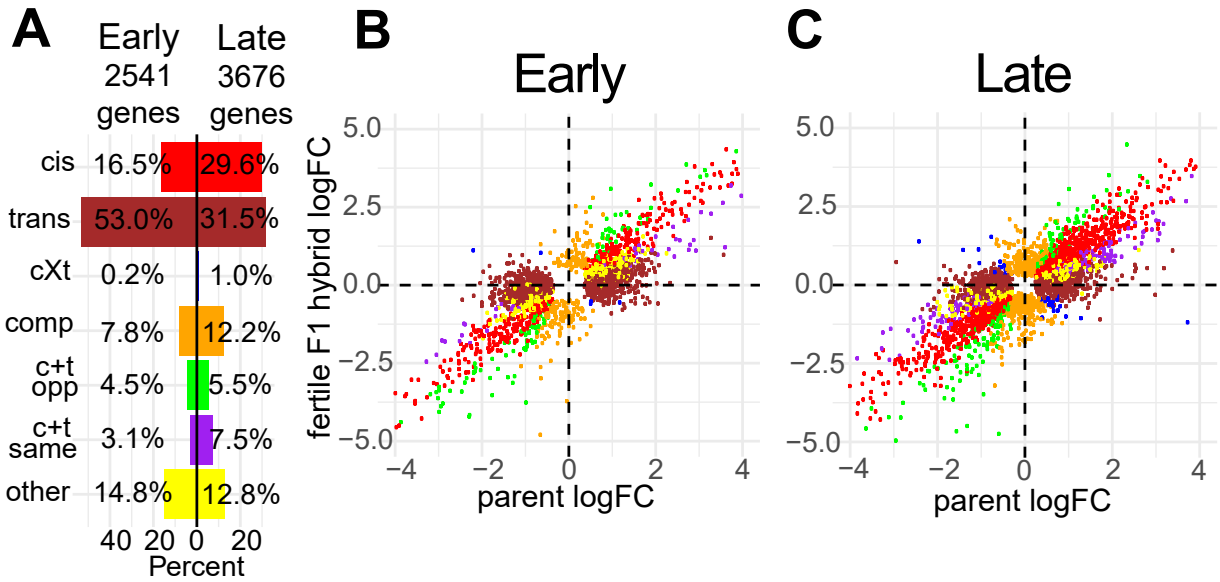

Figure 5. Regulatory category results for the fertile F1 hybrid (LEWES♀ X PWK♂) (*A*) Percent of non-conserved genes in each regulatory category both early and late. (*B* and *C*) Expression logFC between alleles within the fertile F1 (y-axis) plotted against the expression logFC between the parental subspecies (x-axis). Each point represents a single gene. Colors correspond to (*A*) and indicate the regulatory category to which that gene was assigned. cXt = *cis* X *trans*; comp = compensatory; c+t opp = *cis* + *trans* opposite; c+t same = *cis* + *trans* same

more common during early spermatogenesis (Figure 5), and that expression level tended to be more conserved early (Figure 3), we hypothesized that *trans*-regulatory changes would have smaller effect sizes, consistent with some previous studies (Coolon, et al. 2014; Metzger, et al. 2016; Hill, et al. 2020). Consistent with this hypothesis, *trans* changes had lower median divergence than *cis* changes based on both EVE divergence and expression logFC (Figure 6). We saw higher divergence for reinforcing mutations based on logFC, but not based on EVE (Figure 6). Our regulatory category analyses were based only on *dom* and *mus*, so it may be that many reinforcing changes are specific to this comparison and cause high divergence between the two subspecies, but these same genes may not accumulate more divergence at deeper phylogenetic levels. For the early cell type, 26% of genes in the reinforcing category overlapped with genes that had high pairwise divergence between *dom* and *mus*, whereas only 10-16% of genes in this category overlapped with high divergence genes in other pairwise comparisons (supplementary table S6). Similarly, 21% of late cell type genes in the reinforcing category overlapped with highly divergent *mus* vs *dom* genes, but only 9-14% of genes in this category overlapped with other pairwise comparisons (supplementary table S6). Collectively, our results showed that *cis*-regulatory changes tended to have larger effects on expression divergence than *trans*-regulatory changes, and that reinforcing mutations tended to have large effects on expression divergence in pairwise comparisons between *mus* and *dom*, but not at deeper levels of evolutionary divergence.

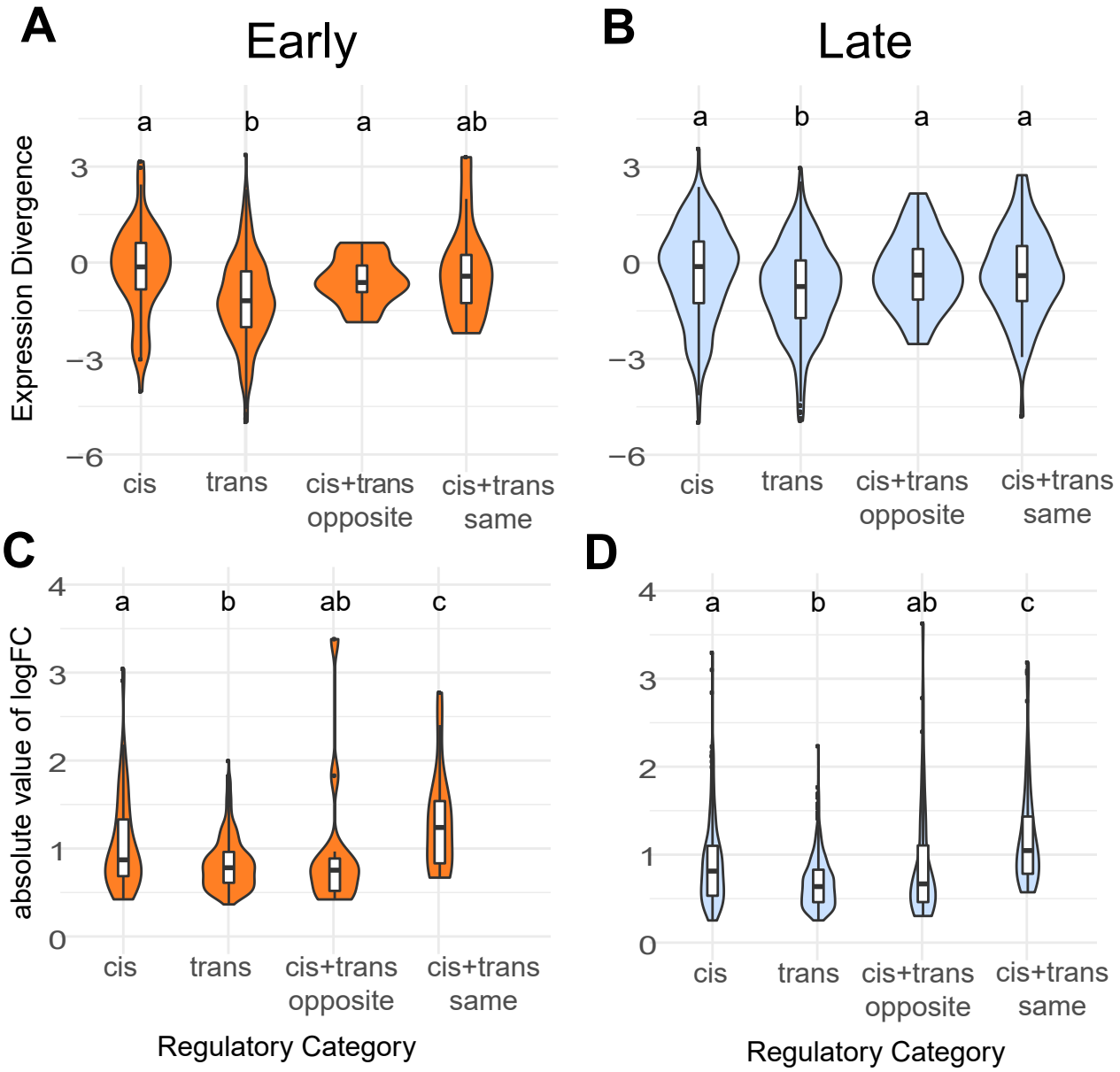

Figure 6. Expression divergence violin plots by regulatory category for the fertile hybrid. Expression divergence is calculated using the value from EVE (*A* and *B*) and as the absolute value of the logFC in expression between parental subspecies (*C* and *D*). Plots *A* and *C* correspond to the early cell type and plots *B* and *D* correspond to the late cell type. Letters indicate significant differences between categories based on a pairwise Wilcoxon rank sum test.

Discussion

Developmental stage and context play an important role in shaping the molecular evolution of reproductive genes (Dean, et al. 2009; Larson, et al. 2016; Finseth and Harrison 2018; Schumacher and Herlyn 2018), with genes expressed in later developmental stages evolving more rapidly (Good and Nachman 2005; Larson, et al. 2016). However, comparing gene expression and protein divergence across developmental stages has rarely been done in a phylogenetic framework. In this study, we combined comparative genomics with cell sorting in four species to understand mouse spermatogenesis evolution across a common developmental framework. Our results give insight into how evolution proceeds at different stages of sperm development, at different molecular levels, and on different chromosome types.

Molecular Divergence across Development

There is a long-standing prediction that early developmental stages should be more constrained, with evolutionary divergence gradually increasing across development ("funnel" model or von Baer's Law, reviewed in Abzhanov 2013). In contrast, many studies have found more rapid gene expression divergence during both the early and late stages of development, with the middle stages showing the most conservation ("hourglass" model; Domazet-Lošo and Tautz 2010; Kalinka, et al. 2010). It remains unclear if the funnel or hourglass model represents a more accurate view of evolution across developmental processes (Abzhanov 2013; Piasecka, et al. 2013), and we cannot address this directly with our dataset that includes two timepoints.
In addition to relaxed developmental constraint in late spermatogenesis, these postmeiotic stages are enriched for genes with narrower expression profiles or highly specific biological functions and are therefore expected to experience relaxed pleiotropic constraint (Eddy 2002; Good and Nachman 2005; Green, et al. 2018). All these models predict higher divergence in the later stages of development, motivating our general hypothesis that the postmeiotic round spermatid stage would diverge more rapidly than the early meiosis spermatocyte stage.

Sexual selection is also likely to be a primary determinant of spermatogenic evolution, and variation in the intensity of sexual selection across spermatogenesis is not well understood (White-Cooper, et al. 2009). Sperm competition and cryptic female choice can select for changes in sperm production rate, form, or function (Pizzari and Parker 2009), and many aspects of sperm morphology correlate with the intensity of post-mating sexual selection (Firman and Simmons 2008; Pitnick, et al. 2009; Lüpold, et al. 2016; McLennan, et al. 2017; Pahl, et al. 2018). Rates of mitotic and initial meiotic divisions during early spermatogenesis can control the overall rate of sperm production (Ramm and Schärer 2014). Therefore, selection for increased sperm production likely acts during the development of spermatogonia, the diploid mitotic cells (White-Cooper, et al. 2009). In contrast, sexual selection shaping the form and function of mature sperm (e.g., sperm swimming speed and fertilization ability) likely acts on later developmental stages such as haploid spermatids (Alavioon, et al. 2017). However, many genes involved in mature spermatozoa functions are also highly expressed in early meiosis (da Cruz, et al. 2016), suggesting that spermatozoa may be shaped by regulatory networks operating throughout spermatogenesis.

All components of molecular evolution that we considered showed more divergence when considering genes induced in late spermatogenesis: lineage-specific expression (Figure 2), protein-coding divergence, and expression level divergence (Figure 3). On first principles, these likely result from a combination of positive selection and relaxed developmental and pleiotropic constraint (Eddy 2002; Swanson and Vacquier 2002; Winter, et al. 2004; Good and Nachman 2005; Abzhanov 2013; Green, et al. 2018). However, our study was underpowered to detect selection using likelihood ratio test approaches with only four lineages (Anisimova, et al. 2001; Rohlfs and Nielsen 2015). Thus, the relative contributions of positive selection and relaxed constraint to rapid spermatogenesis evolution remain unclear, especially for gene expression phenotypes for which it is particularly difficult to tease apart the relative contributions of positive selection and relaxed constraint.

Our set of induced genes provided strong evidence for rapid evolution late, but results were less clear when looking at different sets of genes. Spermatogenesis is a transcriptionally complex process, with most of the genome expressed in the testes (Soumillon, et al. 2013) and high overlap between genes expressed early and late in our dataset (Table 1). For protein-coding divergence, we saw more rapid evolution late only when looking at the induced dataset, but not when looking at all expressed genes, likely because most genes in our dataset were expressed in both cell types. For expression divergence, there was more rapid evolution late even when looking at all expressed genes. This suggests that even genes with broader (i.e., non-induced) expression patterns have conserved expression levels in early spermatogenesis, but more divergent expression patterns in late spermatogenesis.

We found that testis-specific genes tended to be both induced late and evolving rapidly at the protein-coding level. Testis-specific and male-biased genes often evolve rapidly, which could be the result of positive selection on genes with specific spermatogenesis functions or relaxed pleiotropic constraint because these genes have highly specific functions (Meiklejohn, et al. 2003; Baines, et al. 2008; Meisel 2011; Parsch and Ellegren 2013; Dean and Mank 2016). However, we did not see a significant faster late pattern for protein-coding or pairwise expression divergence when looking only at testis-specific genes. Although there were relatively few genes in the testis-specific set, it appears that testis-specific genes were rapidly evolving regardless of which spermatogenesis stage they are expressed in. If generally true, then our observation of more rapid divergence late in spermatogenesis may partially reflect a higher proportion of testis-specific genes induced in the late cell type (Table 1).

In addition to these broad patterns of molecular evolution, we explored the potential functional relevance of rapid divergence for specific genes. We detected 20

genes with high (>2.5) EVE divergence in either cell type, and of these 15 were broadly expressed, but five may have specific roles in spermatogenesis (The UniProt Consortium 2020). For example, *Rnf19a* had an EVE value of 4.2 in the late cell type and has a known role in the formation of the sex body, which isolates the sex chromosomes in the nucleus during meiosis, a process that is required for proper spermatogenesis (Párraga and del Mazo 2000) and appears to be disrupted in sterile hybrid mice (Bhattacharyya, et al. 2013). Two genes that were rapidly evolving at the protein-coding level are thought to be essential for spermatogenesis: *Ttc29*, which is required for proper sperm tail formation in humans and mice (Liu, et al. 2019; Lorès, et al. 2019), and *Pmis2*, which is required for male fertility in mice (Yamaguchi, et al. 2012).

Gene Expression versus Protein-Coding Divergence

Protein-coding changes alter a gene in every tissue and developmental stage in which it is expressed, whereas expression changes have the potential to be more specific (Wray, et al. 2003; Wray 2007; Carroll 2008). Expression changes, specifically *cis*-regulatory changes, should be less constrained by pleiotropy and may underlie evolutionary changes when purifying selection acts more strongly against protein-coding divergence (Wray, et al. 2003; Carroll 2008). Under this model, we might expect to see less pronounced differences in relative expression levels when comparing early versus late stages. However, more recent work has shown that *cis*-regulatory elements such as enhancers can be highly pleiotropic, so *cis*-regulatory changes may be more constrained than once thought (Sabarís, et al. 2019; Hill, et al. 2020). If gene expression and protein-coding are subject to similar developmental and pleiotropic constraints, we would expect them to show similar evolutionary patterns across spermatogenesis, as we observed for autosomal genes (early versus late spermatogenic cell types, Figure 3).

Interestingly, despite parallel trends in relative divergence across spermatogenesis, expression level divergence and protein-coding divergence were not strongly correlated, suggesting that these two types of molecular changes mostly evolve independently at the individual gene level (Khaitovich, et al. 2005). Perhaps surprisingly, there was no overlap between genes with very rapid protein-coding divergence (dN/dS > 1.5) and high EVE expression divergence (EVE divergence > 2.5). Likewise, only 26 genes with high pairwise expression divergence in at least one comparison (pairwise divergence metric > 1) also had high protein-coding divergence (dN/dS > 1.5; supplementary table S7). Whether expression or protein-coding is more rapid for a particular gene may depend on factors such as expression breadth and protein function, but rarely do spermatogenic genes appear to be rapidly evolving for both gene expression and protein sequences.

We also investigated lineage-specificity, because testes and sperm tend to be enriched for lineage-specific genes (Brawand, et al. 2011; Ahmed-Braimah, et al. 2017; Whittington, et al. 2017) and novel genes (Betrán, et al. 2002; Assis and Bachtrog 2013; Schroeder, et al. 2019; Cridland, et al. 2020; Lange, et al. 2021). Lineage-specific and novel genes may be common in spermatogenesis because testes are highly transcriptionally active and have a high tissue-specific expression profile, which may allow new genes to arise without disrupting other tissues or processes (Levine, et al. 2006; Kaessmann 2010; Soumillon, et al. 2013; Zhao, et al. 2014). We found that late spermatogenesis has a higher proportion of lineage-specific genes than early spermatogenesis. This is likely due to relaxed developmental constraint in the later stages, where novel expression patterns have less potential to disrupt developmental gene regulatory networks and therefore disrupt spermatogenesis. Increased lineage-specificity late is consistent with higher expression level divergence late, because both results suggest that spermatogenesis can tolerate more changes to expression patterns in the late stages without causing sterility. Lineage-specific genes are expected to be less conserved at the protein-coding level, because their novel expression patterns suggest they are not essential for fertility across the mouse phylogeny and therefore less likely to experience purifying selection. Thus, more lineage-specificity late is also consistent with our observation of increased protein-coding divergence late.

X Chromosome Evolution

The X chromosome is predicted to evolve faster than the autosomes because it has a lower effective population size and is therefore more susceptible to drift, and because it is hemizygous in males so beneficial recessive mutations will fix more quickly (Charlesworth, et al. 1987; Vicoso and Charlesworth 2009). Empirical studies show evidence for this effect at the protein-coding level in many taxa, particularly for male reproductive genes (Khaitovich, et al. 2005; Baines, et al. 2008; Meisel and Connallon 2013; Parsch and Ellegren 2013; Larson, et al. 2016), but faster-X evolution is not always observed (Whittle, et al. 2020). Our data provide strong evidence for faster-X protein-coding evolution in both early and late spermatogenesis, suggesting that the faster-X effect applies across genes involved in different spermatogenesis stages in mice.

Our results were more complex for expression evolution. We used both a phylogenetic approach (Rohlfs and Nielsen 2015) and pairwise approach (Meisel, et al. 2012) to analyze expression level divergence, and these two methods sometimes gave contrasting results. In the early cell type, pairwise comparisons gave evidence for a faster-X effect for gene expression, while the phylogenetic model did not (Figure 3*B*, supplementary table S3). In the early cell type, correlations between different pairwise divergence values were relatively low on the X chromosome, suggesting that X-linked genes with high expression level divergence in one pairwise comparison did not tend to have high divergence in other comparisons (Table 2). It is well-established that lineage-specific changes can create false signatures of rapid divergence in pairwise comparisons (Felsenstein 1985), and a recent study found this issue in studies of gene expression evolution (Dunn, et al. 2018). These contrasting results in the early cell type highlight the importance of evaluating molecular evolution in a phylogenetic context, because this approach controls for shared evolutionary history among lineages and gives

insight into more general evolutionary trends (Rohlfs and Nielsen 2015; Dunn, et al. 2018). In the late cell type, both phylogenetic and pairwise divergence metrics support a similar rate of X-linked and autosomal expression evolution (Figure 3*B*, supplementary table S3).

Overall, our results do not support a faster-X effect for gene expression evolution, in contrast to several previous studies (Khaitovich, et al. 2005; Voolstra, et al. 2007; Brawand, et al. 2011; Meisel, et al. 2012; Coolon, et al. 2015). These studies were in other systems and used whole testes samples, which are made up of different cell types, so signals of expression divergence may partially reflect differences in cell type composition rather than true per cell changes in expression levels (Wistuba, et al. 2003; Good, et al. 2010; Cridland, et al. 2020; Hunnicutt, et al. 2021; Yapar, et al. 2021). One previous study used cell-type specific data and showed that the X chromosome is under-enriched for differentially expressed genes during late spermatogenesis in a pairwise comparison of two different mouse lineages (Larson, et al. 2016), and our phylogenetic sampling demonstrates that this result likely applies across mice.

Contrasting patterns of expression level and protein sequence evolution on the X chromosome during spermatogenesis suggest that these two levels of molecular evolution experience different evolutionary pressures (Larson, et al. 2016). Theoretical predictions for the faster-X effect on protein-coding evolution may also apply to gene expression changes, but only for *cis*-regulatory changes where the causative mutations are on the X chromosome (Meisel and Connallon 2013; Larson, et al. 2016). During spermatogenesis, the sex chromosomes undergo MSCI (McKee and Handel 1993; Turner 2015) and PSCR (Namekawa, et al. 2006; Sin, et al. 2012; Larson, et al. 2016), which likely imposes an overall repressive regulatory environment that constrains gene expression levels but not protein-coding changes. Disruption of MSCI and PSCR strongly impairs male fertility, so evolutionary constraints on X chromosome expression during spermatogenesis are expected to be strong (Burgoyne, et al. 2009; Good, et al. 2010; Larson, et al. 2017). Our results support the hypothesis that unique regulatory constraints on the X chromosome during spermatogenesis reduce expression level divergence, while still allowing rapid protein-coding divergence (Larson, et al. 2016; Larson, et al. 2018a). This finding underscores the importance of investigating molecular evolution at multiple levels, because different components of molecular evolution may experience unique evolutionary pressures that result in distinct patterns of divergence (Brawand, et al. 2011; Halligan, et al. 2013; Larson, et al. 2016).

Regulatory Mechanisms Underlying Expression Divergence

Understanding the relative contribution of *cis*- versus *trans*-acting mutations underlying expression divergence remains an important and largely unanswered question in molecular evolution (Benowitz, et al. 2020; Hill, et al. 2020). Resolving the relative contributions of these regulatory mechanisms is an important step towards understanding the genetic architecture of expression phenotypes and how different evolutionary forces may act on gene expression (Hill, et al. 2020). Although considerable progress has been made in a few key model systems (Goncalves, et al. 2012; Coolon, et al. 2014; Mack, et al. 2016; Metzger, et al. 2016; Combs and Fraser 2018; Benowitz, et al. 2020; Cridland, et al. 2020; Hill, et al. 2020; Sánchez-Ramírez, et al. 2021), available data on *cis*- versus *trans*-regulatory changes mostly come from whole tissues or organisms. Our results showed that the relative contribution of underlying regulatory mechanisms can differ dramatically between two cell types within a single complex tissue. While these striking differences are perhaps an expected consequence of different selective pressures acting on cellular function and developmental stage, they also underscore how underlying regulatory phenotypes might be difficult to resolve from complex tissues.

As *cis*-regulatory changes are generally thought to be less pleiotropic (Wray 2007; Carroll 2008; Hill, et al. 2020), we predicted that *cis*-regulatory mutations would be proportionally more common in early spermatogenesis compared to late spermatogenesis. However, we found the opposite pattern with more *cis*-regulatory changes late compared to early (Figure 5, supplementary table S5). The relative contributions of *cis*- and *trans*-regulatory changes to expression divergence likely result from multiple factors in addition to differences in pleiotropic constraint, including a larger mutational target size for *trans* mutations and a tendency of *cis* mutations to have larger effect sizes (Coolon, et al. 2014; Metzger, et al. 2016; Hill, et al. 2020). While our observation of proportionally more *cis*-regulatory changes late was inconsistent with our original prediction, it was consistent with a larger effect size for cis-regulatory mutations (Figure 6*C*, 6*D*) and higher overall expression divergence late (Figure 3). Thus, individual effect sizes may play a more important role than pleiotropic constraint in determining the relative contribution of different regulatory mechanisms across spermatogenesis. There is also increasing evidence that the relative pleiotropy of *cis*- and *trans*-mutations is more complicated than once proposed, as many *cis*-acting enhancers appear to be highly pleiotropic (Sabarís, et al. 2019; Hill, et al. 2020).

Cis- and *trans*-regulatory mutations can combine to affect the expression of a single gene, either in the same direction (reinforcing) or in opposite directions (compensatory; Goncalves, et al. 2012; Coolon, et al. 2014; Mack, et al. 2016). We observed a higher proportion of compensatory mutations than reinforcing mutations across both spermatogenesis cell types and in whole testes. This was expected given that gene expression tends to evolve under stabilizing selection (Rohlfs and Nielsen 2015), and it is consistent with previous studies across many tissue types in mice (Goncalves, et al. 2012; Mack, et al. 2016), flies (Coolon, et al. 2014; Benowitz, et al. 2020), and roundworms (Sánchez-Ramírez, et al. 2021). We also saw relatively more reinforcing mutations during postmeiotic spermatogenesis. Reinforcing mutations tended to have a larger effect size based on expression logFC between *mus* and *dom* (Figure 6*C*, 6*D*), so our observation of more large-effect reinforcing changes late is consistent with higher

expression level divergence in late spermatogenesis.

Given the striking differences that we saw between just two cell types, it is likely that complex tissues composed of many cell types may often give different results than isolated cell populations. Consistent with this prediction, our observed proportions of genes in each regulatory category differ from some other published results in house mouse whole tissues (i.e., liver, Goncalves, et al. 2012; whole testes, Mack, et al. 2016), primarily in that we saw a much higher proportion of genes in the *trans* category. We also found some different patterns when reanalyzing whole testes expression data from Mack, et al. (2016) that likely reflect technical differences in the analytical pipelines used between studies (supplementary table S5, see Supplementary Methods for details). In general, our pipeline used comparatively more conservative approaches to test for significant DE or ASE. Thus, only genes showing relatively pronounced differences in expression levels between genotypes or alleles were assigned to regulatory mechanisms in our study.

We also found that the relative proportion of *cis*- and *trans*-regulatory changes were similar between whole testes and the late cell type in the fertile F1 hybrid (supplementary table S5), consistent with the observation that postmeiotic spermatids have a disproportionately large contribution to mouse whole testes expression patterns (Hunnicutt, et al. 2021). These results suggest that changes in the relative intensities of evolutionary pressures acting across spermatogenesis not only change the relative rates of expression level divergence, but also select for different mechanisms of regulatory evolution underlying these expression changes. Given this, analyzing such patterns at the level of whole organisms or tissues is unlikely to provide a clear understanding of how mechanisms of regulatory evolution proceed in underlying cells. Indeed, even enriched cell populations as we have generated may be limited by relative purities.

By considering both expression divergence across the *Mus* phylogeny and underlying mechanisms of regulatory divergence between two lineages (*mus* and *dom*), our study also provides an opportunity to connect different types of regulatory changes to patterns of expression divergence at a deeper phylogenetic scale. Although *trans*-acting changes were relatively common (Figure 5), genes with *cis*-regulatory changes between *mus* and *dom* tended to have higher phylogeny-wide expression divergence than those with *trans*-regulatory changes for both cell types (Figure 6*A*, 6*B*). This suggests that genes showing *cis*-regulatory changes are more likely to accumulate regulatory differences over time, resulting in phylogeny-wide expression divergence, whereas genes showing *trans*-regulatory changes at relatively shallow evolutionary scales tend to be relatively conserved across the *Mus* phylogeny. Genes with reinforcing changes also had relatively low phylogeny-wide expression level divergence (Figure 6*A*, 6*B*), in contrast to their high pairwise divergence between *mus* and *dom* (Figure 6*C*, 6*D*). Genes in this category likely have large-effect, lineage-specific changes in expression that may be under purifying selection over deeper phylogenetic levels. Finally, our phylogenetic contrast revealed rapid expression level divergence late in spermatogenesis. By combining these data with allele-specific expression data, we further showed that *cis*-regulatory changes are likely to underlie this rapid phylogeny-wide expression divergence in late spermatogenesis.

Materials and Methods

Mouse Resources

We investigated gene expression and protein-coding evolution in 12 *Mus musculus domesticus* (*dom*) individuals from four inbred strains (2 BIK/g, 3 DGA, 3 LEWES/EiJ, 4 WSB/EiJ), 8 *M. m. musculus* (*mus*) individuals from three inbred strains (2 CZII/EiJ, 3 MBS, 3 PWK/PhJ), 11 *M. spretus* (*spr*) individuals from three inbred strains (5 SEG, 2 SFM, 4 STF), and 3 *M. pahari* (*pah*) individuals from one inbred strain (3 PAHARI/EiJ; Figure 1*B*). By using multiple wild-derived inbred strains of *dom*, *mus*, and *spr*, we sampled natural within-species variation while also having biological replicates of genetically similar individuals. These mice were maintained in breeding colonies at the University of Montana (UM) Department of Laboratory Animal Resources (IACUC protocol 002-13). These colonies were initially established from mice purchased from The Jackson Laboratory, Bar Harbor, ME (CZECHII/EiJ, PWK/PhJ, WSB/EiJ, LEWES/EiJ, PAHARI/EiJ) or acquired from Matthew Dean's colonies at the University of Southern California which were derived from François Bonhomme's stocks at the University of Montpellier, Montpellier, France (MBS, BIK, DGA, STF, SFM, SEG). We weaned males at ~21 days postpartum (dpp) into same sex sibling groups and we caged males individually at least 15 days prior to euthanization to avoid dominance effects on testes expression. We euthanized mice at 60-160 dpp by CO2 followed by cervical dislocation.

For expression data from reciprocal F1 males, we used FACS enriched expression data from (Larson, et al. 2017). These data include males from reciprocal F1 crosses between different inbred strains within each *M. musculus* subspecies (*mus*: CZECHII females X PWK males, *dom*: WSB females X LEWES males), as well as reciprocal *mus* and *dom* F1 hybrids (LEWES females X PWK males and PWK females X LEWES males), allowing us to compare results at two different levels of divergence (i.e., within and between lineages). We also analyzed whole testes expression data from (Mack, et al. 2016) to compare FACS enriched cell-types to whole testes, including crosses between different strains within each *M. musculus* subspecies (LEWES females X WSB males and PWK females X CZII males) and the same reciprocal F1 hybrid crosses to those in (Larson, et al. 2017).

Testis Cell Sorting and RNAseq

We collected testes from mice immediately following euthanization and isolated cells at different stages of spermatogenesis using Fluorescence Activated Cell Sorting (FACS) as previously described (Larson, et al. 2016). Briefly, we decapsulated testes and washed them with twice with 1mg/mL collagenase (Worthington

Biochemical), 0.004mg/mL DNase I (Qiagen), and GBSS (Sigma), followed by disassociation with 1mg/mL trypsin (Worthington Biochemical) and 0.004mg/mL DNase I. We then inactivated trypsin with 0.16mg/mL fetal calf serum (Sigma). For each wash and disassociation step, we incubated and agitated samples at 33°C for 15 minutes on a VWR minishaker at 120 rpm. We stained cells with 0.36mg/mL Hoechst 33324 (Invitrogen) and 0.002mg/mL propidium iodide, filtered with a 40µm cell filter, and sorted using a FACSAria IIu cell sorter (BD Biosciences) at the UM Center for Environmental Health Sciences Fluorescence Cytometry Core. We periodically added approximately 0.004mg/mL DNase I as needed during sorting to prevent DNA clumps from clogging the sorter. We sorted cells into 15µL beta-mercaptoethanol (Sigma) per 1mL of RLT lysis buffer (Qiagen) and kept samples on ice whenever they were not in the incubator or the cell sorter. For this study, we focused on two cell populations: early meiotic spermatocytes (leptotene/zygotene) and postmeiotic round spermatids. The full FACS experimental protocol, with slight modifications from (Getun, et al. 2011), is available on Github (https://github.com/goodest-goodlab/good-protocols/tree/main/protocols/FACS).

We extracted RNA from sorted cell populations using the Qiagen RNeasy Blood and Tissue Kit and analyzed RNA quality using RNA integrity numbers (RIN) with a Bioanalyzer 2000 (Agilent) or TapeStation 2200 (Agilent). Except for one partially degraded sample (RIN 6.9), all sequenced samples had RIN ≥ 7 and 57 out of 68 samples had RIN > 8 (supplementary table S8). We prepared RNAseq libraries using the Agilent SureSelect protocol and sequenced samples at the Hudson Alpha Institute for Biotechnology using Illumina NextSeq (75bp single end).

Mus strain phylogeny

We generated the phylogeny in Figure 1*B* using available exome (Chang, et al. 2017; Sarver, et al. 2017) or whole genome (Keane, et al. 2011; Thybert, et al. 2018) sequence data (PRJNA326865, PRJNA323493, PRJEB2003, PRJEB14896). Genotypes were based on iterative mapping assemblies relative to the house mouse reference genome (mm10) conducted using pseudo-it v3.0 (Sarver, et al. 2017) that restricts genotyping to user specified genomic regions corresponding to targeted exons. We ran pseudo-it with one iteration to map reads, call variants, and generate consensus fasta files for each sample. We then extracted exons, aligned these regions using MAFFT v7.271 (Katoh and Standley 2013), converted to PHYLIP format using AMAS (Borowiec 2016), and inferred a maximum likelihood concatenated tree using IQ-TREE v2.1.4-beta (Nguyen, et al. 2014).

Processing and Quantification of Gene Expression Data

We used R version 3.6.3 and Bioconductor version 3.10 for all analyses. We trimmed raw reads for adaptors and low-quality bases using expHTS (Streett, et al. 2015) and mapped trimmed reads with TopHat version 2.1.0 (Kim, et al. 2013). Publicly available genome assemblies were available for all four lineages (Keane, et al. 2011; Thybert, et al. 2018), allowing us to map reads to the correct assembly and reduce reference bias (Sarver, et al. 2017). Mapping rates were generally consistent across lineages (supplementary table S8). To select orthologous genes among the four lineages for all subsequent analyses, we used BiomaRt (Durinck, et al. 2005; Durinck, et al. 2009) to identify one-to-one Ensembl orthologs and retain only those that were present in all genome assemblies and the mouse reference build GRCm38.

We used edgeR 3.28.1 (Robinson, et al. 2010) to normalize expression data, calculate fragments per kilobase per million reads (FPKM), and perform differential expression (DE) analyses. A gene was defined as "expressed" in our dataset if it had an FPKM > 1 in at least eight samples. A gene was expressed in a particular lineage and cell type if it had an FPKM > 1 in all samples of that lineage and cell type. A gene was considered induced in a particular cell type if its median FPKM in that cell type across all lineages was greater than two times its median FPKM in the other cell type across all lineages. Testis-specific genes were those only expressed in testis based on the mouse tissue expression data from (Chalmel, et al. 2007).

We defined lineage-specific genes in two ways. First, we used a log fold change (logFC) method in which a gene was considered lineage-specific if its median expression level in a lineage was greater than two times its median expression level in any of the other three lineages. Second, we used a Bayesian approach to determine if a gene is active or inactive in an expression dataset based on transcript levels as implemented with the program Zigzag (Thompson, et al. 2020). Genes identified as being active (posterior P > 0.5) in one lineage and inactive (posterior P < 0.5) in the other lineages were considered lineage-specific. We ran Zigzag twice and only included genes with consistent active or inactive assignments between the two runs. Both the logFC and Zigzag analyses were performed for each cell type, so a gene could be lineage-specific in one cell type but not the other. For each lineage, we determined the proportion of expressed (logFC) or active (Zigzag) genes that were lineage-specific and used a Pearson's $\chi 2$ test to determine if one cell type had greater lineage-specificity than the other.

Protein-Coding Divergence

We used the "iqtree-omp" command in IQTree version 1.5.5 (Nguyen, et al. 2014) to infer a mouse species tree based on gene trees estimated from the reference sequences for all four mouse lineages (Keane, et al. 2011; Thybert, et al. 2018). We took the longest transcript for all one-to-one orthologs and aligned these using MAFFT v7.271 (Katoh and Standley 2013) and converted to PHYLIP format using AMAS (Borowiec 2016). We used a custom script to exclude genes that did not begin with a start codon, had early stop codons, or had sequence lengths that were not multiples of three.

We then used the Codeml program in the PAML package to calculate protein-coding divergence and test for positive selection on protein-coding genes (Yang 2007). We used the M0 model to calculate phylogeny-wide dN/dS for

each gene, which we report as the overall protein-coding divergence values. We also performed a likelihood ratio test between the M8 and M8a site-based models to test for positive directional selection on each gene (Swanson, et al. 2003).

Differential Expression

We performed all analyses of expression level divergence for three different gene sets: expressed genes, induced genes, and testis-specific genes. To calculate expression divergence in a phylogenetic framework, we used the Expression Variance and Evolution (EVE) model (Rohlfs and Nielsen 2015), which performs a phylogenetic ANOVA using an Ornstein-Uhlenbeck model to evaluate divergence while controlling for evolutionary relatedness. The Ornstein-Uhlenbeck model is similar to the Brownian motion model, but with traits being pulled towards an optimal value consistent with stabilizing selection, which is thought to be a pervasive force acting on gene expression (Rohlfs and Nielsen 2015). Although EVE also provides a likelihood test framework to test for positive selection, we did not have adequate power to perform this test with our sampling of four relatively closely related lineages. We report expression divergence from EVE as -log($beta_i$), where $beta_i$ is a metric from EVE that represents the ratio of within-lineage variance to between-lineage evolutionary divergence. By taking the negative log, higher positive numbers correspond to greater divergence between lineages. We excluded genes with extremely low divergence values [-log($beta_i$)<-5] because this subset did not show a linear relationship between evolutionary divergence and population variance and therefore violated underlying assumptions of the EVE model (supplementary figure S6).

We also calculated expression divergence in a pairwise framework (Meisel, et al. 2012). This method takes the difference in expression level between two lineages and normalizes based on the average expression of the gene in both lineages:

$$D_{a,ij} = \frac{|S_{a,i} - S_{a,j}|}{(S_{a,i} + S_{a,j}) / 2}$$

$D_{a,ij}$ is the divergence of gene *a* between lineages *i* and *j*. $S_{a,i}$ is the median FPKM of gene *a* in lineage *i,* and $S_{a,j}$ is the median FPKM of gene *a* in lineage *j*. We also calculated the logFC in expression between every pairwise comparison of lineages as an additional pairwise divergence metric (Robinson, et al. 2010). For the EVE, pairwise divergence, and logFC methods, we compared relative expression divergence between cell types and between the X chromosome and autosomes using a Wilcoxon rank-sum test. We tested if certain cell types or chromosome types show greater correlation among pairwise divergence values using Spearman's rank correlation.

We also compared relative expression divergence on the X chromosome versus the autosomes using the proportion of DE genes on each chromosome (Good, et al. 2010; Larson, et al. 2016). First, we calculated the proportion of expressed genes that are DE across all autosomes. We then multiplied this proportion by the number of genes expressed on each chromosome to calculate the expected number of DE genes for each chromosome. We plotted the observed number of DE genes against the expected number and used a hypergeometric test to evaluate if each chromosome is over- or under-enriched for DE genes.

Allele-Specific Expression and Regulatory Divergence

We used the modtools and lapels-suspenders pipeline (Holt, et al. 2013; Huang, et al. 2014) to reduce mapping bias and to assign the parental origin of reads in F1 individuals (See Supplementary Methods). This approach requires mapping to pseudogenomes generated using modtools to resolve differences in genome coordinates between different references. We used publicly available pseudogenomes for WSB and PWK, which incorporate single nucleotide variants (SNVs) and indels from these strains into the GRCm38 mouse reference build (Huang, et al. 2014). For LEWES and CZECHII, we generated our own pseudogenomes with modtools version 1.0.2 using published LEWES (Morgan, et al. 2016) and CZECHII (Larson, et al. 2018b) VCF files to identify SNVs and indels for these strains. We developed a custom pipeline (See Supplementary Methods) to assign genes to regulatory categories following previous recommendations (Coolon, et al. 2014; Mack, et al. 2016; Combs and Fraser 2018; Benowitz, et al. 2020). To determine significant differences between cell types, we performed a Pearson's $\chi 2$ test followed by false discovery rate correction for multiple tests.

Competing Interests

The authors declare they have no conflict of interest.

Data Availability

RNAseq data generated for this project are available through the National Center for Biotechnology Information under accession PRJNA735780. Individual sample accessions are in supplementary table S8. A table of genes included in our analyses and whether they were considered expressed, induced, or active in each cell type is available in supplementary table S9.

Author contributions

J.M.G conceived and funded the project. E.E.K.K., E.L.L., and J.M.G. designed the experiments. E.L.L. and S.K. performed the mouse husbandry and breeding experiments. C.C., S.K., and E.L.L. performed mouse dissections and cell sorts. C.C. and E.L.L. prepared the sequencing libraries. E.E.K.K. analyzed the data. E.E.K.K., E.L.L., and J.M.G. wrote the manuscript with input from all authors.

Acknowledgements

We would like to thank Pamela K. Shaw and the UM Fluorescence Cytometry Core supported by an Institutional Development Award from the NIGMS (P30GM103338), the UM Genomics Core supported by the M.J. Murdock Charitable Trust, the UM Lab Animal Resources staff, Gregg Thomas for assistance generating the phylogeny in Figure

1*B*, Nathanael Herrera for mouse photos, and the Good Lab for helpful advice. This work was supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (R01-HD073439, R01-HD094787 to JMG). EEKK was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1313190). ELL was supported by the National Science Foundation (DEB-2012041). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

References

Abzhanov A. 2013. von Baer's law for the ages: lost and found principles of developmental evolution. Trends Genet. 29:712-722.

Ahmed-Braimah YH, Unckless RL, Clark AG. 2017. Evolutionary Dynamics of Male Reproductive Genes in the *Drosophila virilis* Subgroup. G3-Genes Genom Genet. 7:3145-3155.

Alavioon G, Hotzy C, Nakhro K, Rudolf S, Scofield DG, Zajitschek S, Maklakov AA, Immler S. 2017. Haploid selection within a single ejaculate increases offspring fitness. Proc Natl Acad Sci U S A. 114:8053-8058.

Albrechtová J, Albrecht T, Ďureje L, Pallazola VA, Piálek J. 2014. Sperm Morphology in Two House Mouse Subspecies: Do Wild-Derived Strains and Wild Mice Tell the Same Story? PLoS One. 9:e115669.

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. Mol Biol Evol. 18:1585-1592.

Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. Proc Natl Acad Sci U S A. 110:17409-17414.

Baines JF, Harr B. 2007. Reduced X-Linked Diversity in Derived Populations of House Mice. Genetics. 175:1911-1921.

Baines JF, Sawyer SA, Hartl DL, Parsch J. 2008. Effects of X-Linkage and Sex-Biased Gene Expression on the Rate of Adaptive Protein Evolution in *Drosophila*. Mol Biol Evol. 25:1639-1650.

Benowitz KM, Coleman JM, Allan CW, Matzkin LM. 2020. Contributions of *cis*- and *trans*-Regulatory Evolution to Transcriptomic Divergence across Populations in the *Drosophila mojavensis* Larval Brain. Genome Biol Evol. 12:1407-1418.

Betrán E, Thornton K, Long M. 2002. Retroposed New Genes Out of the X in *Drosophila*. Genome Res. 12:1854-1859.

Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P, Simecek P, Forejt J. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. Proc Natl Acad Sci U S A. 110:E468-E477.

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 4:e1660.

Brawand D, Soumillon M, Necsulea A, Julien P, Csardi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. Nature. 478:343-348.

Burgoyne PS, Mahadevaiah SK, Turner JMA. 2009. The consequences of asynapsis for mammalian meiosis. Nat Rev Genet. 10:207-216.

Cao J, Zhou W, Steemers F, Trapnell C, Shendure J. 2020. Sci-fate characterizes the dynamics of gene expression in single cells. Nat Biotechnol. 38:980-988.

Carroll SB. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell. 134:25-36.

Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, Gattiker A, Moore J, Patard J-J, Wolgemuth DJ, Jégou B, et al. 2007. The conserved transcriptome in human and rodent male gametogenesis. Proc Natl Acad Sci U S A. 104:8346-8351.

Chang PL, Kopania E, Keeble S, Sarver BAJ, Larson E, Orth A, Belkhir K, Boursot P, Bonhomme F, Good JM, et al. 2017. Whole exome sequencing of wild-derived inbred strains of mice improves power to link phenotype and genotype. Mamm Genome. 28:416-425.

Charlesworth B, Coyne JA, Barton NH. 1987. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. Am Nat. 130:113-146.

Chevret P, Veyrunes F, Britton-Davidian J. 2005. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. Biol J Linn Soc. 84:417-427.

Combs PA, Fraser HB. 2018. Spatially varying *cis*-regulatory divergence in *Drosophila* embryos elucidates *cis*-regulatory logic. PLoS Genet. 14:e1007631.

Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. Genome Res. 24:797-808.

Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ. 2015. Molecular Mechanisms and Evolutionary Processes Contributing to Accelerated Divergence of Gene Expression on the *Drosophila* X Chromosome. Mol Biol Evol. 32:2605-2615.

Cridland JM, Majane AC, Sheehy HK, Begun DJ. 2020. Polymorphism and Divergence of Novel Gene Expression Patterns in *Drosophila melanogaster*. Genetics. 216:79-93.

Cutter AD, Bundus JD. 2020. Speciation and the developmental alarm clock. eLife. 9:e56276.

da Cruz I, Rodríguez-Casuriaga R, Santiñaque FF, Farías J, Curti G, Capoano CA, Folle GA, Benavente R, Sotelo-Silveira JR, Geisinger A. 2016. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. BMC Genomics. 17:294.

Dean MD, Ardlie KG, Nachman MW. 2006. The frequency of multiple paternity suggests that sperm competition is common in house mice (*Mus domesticus*). Mol Ecol. 15:4141-4151.

Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW. 2009. Proteomics and Comparative Genomic Investigations Reveal Heterogeneity in Evolutionary Rate of Male Reproductive Proteins in Mice (*Mus domesticus*). Mol Biol Evol. 26:1733-1743.

Dean R, Mank JE. 2016. Tissue Specificity and Sex-Specific Regulatory Variation Permit the Evolution of Sex-Biased Gene Expression. Am Nat. 188:E74-84.

Domazet-Lošo T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature. 468:815-818.

Dunn CW, Zapata F, Munro C, Siebert S, Hejnol A. 2018. Pairwise comparisons across species are problematic when analyzing functional genomic data. Proc Natl Acad Sci U S A. 115:E409-E417.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 21:3439-3440.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 4:1184-1191.

Eddy EM. 2002. Male germ cell gene expression. Recent Prog Horm Res. 57:103-128.

Felsenstein J. 1985. Phylogenies and the Comparative Method. Am Nat. 125:1-15.

Finseth FR, Harrison RG. 2018. Genes Integral to the Reproductive Function of Male Reproductive Tissues Drive Heterogeneity in Evolutionary Rates in Japanese Quail. G3-Genes Genom Genet. 8:39-51.

Firman RC. 2020. Of mice and women: advances in mammalian sperm competition with a focus on the female perspective. Philos Trans R Soc Lond, Ser B: Biol Sci. 375:20200082.

Firman RC, Simmons LW. 2008. Polyandry, sperm competition, and reproductive success in mice. Behav Ecol. 19:695-702.

Getun IV, Torres B, Bois PRJ. 2011. Flow Cytometry Purification of Mouse Meiotic Cells. JoVE-J Vis Exp.2602.

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.

Good JM, Giger T, Dean MD, Nachman MW. 2010. Widespread Over-Expression of the X Chromosome in Sterile F1 Hybrid Mice. PLoS Genet. 6:e1001148.

Good JM, Handel MA, Nachman MW. 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. Evolution. 62:50-65.

Good JM, Nachman MW. 2005. Rates of Protein Evolution Are Positively

Correlated with Developmental Timing of Expression During Mouse Spermatogenesis. Mol Biol Evol. 22:1044-1052.
Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB, et al. 2018. A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Dev Cell.
Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, Adams DJ, Keightley PD. 2013. Contributions of Protein-Coding and Regulatory Change to Adaptive Molecular Evolution in Murid Rodents. PLoS Genet. 9:e1003995.
Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, Mank JE. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. Proc Natl Acad Sci U S A. 112:4393-4398.
Hill MS, Vande Zande P, Wittkopp PJ. 2020. Molecular and evolutionary processes generating variation in gene expression. Nat Rev Genet.
Holt J, Huang S, McMillan L, Wang W. 2013. Read Annotation Pipeline for High-Throughput Sequencing Data. Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics; Washington DC, USA: 2506645: ACM. p. 605-612.
Huang S, Holt J, Kao C-Y, McMillan L, Wang W. 2014. A novel multi-alignment pipeline for high-throughput sequencing data. Database. 2014.
Hunnicutt KE, Good JM, Larson EL. 2021. unpublished data. https://wwwbiorxivorg/content/101101/20210708451646v1. last accessed July 12, 2021.
Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20:1313-1326.
Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model. Nature. 468:811-814.
Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 30:772-780.
Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 477:289-294.
Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005. Parallel Patterns of Evolution in the Genomes and Transcriptomes of Humans and Chimpanzees. Science. 309:1850-1854.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36.
King M, Wilson A. 1975. Evolution at two levels in humans and chimpanzees. Science. 188:107-116.
Kousathanas A, Halligan DL, Keightley PD. 2014. Faster-X Adaptive Protein Evolution in House Mice. Genetics. 196:1131-1143.
Lange A, Patel PH, Heames B, Damry AM, Saenger T, Jackson CJ, Findlay GD, Bornberg-Bauer E. 2021. Structural and functional characterization of a putative de novo gene in *Drosophila*. Nat Commun. 12:1667.
Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein-coding genes in *Drosophila*. Trends Genet. 24:114-123.
Larson EL, Keeble S, Vanderpool D, Dean MD, Good JM. 2017. The Composite Regulatory Basis of the Large X-Effect in Mouse Speciation. Mol Biol Evol. 34:282-295.
Larson EL, Kopania EEK, Good JM. 2018a. Spermatogenesis and the Evolution of Mammalian Sex Chromosomes. Trends Genet. 34:722-732.
Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BAJ, Smith AD, Dean MD, Good JM. 2016. Contrasting Levels of Molecular Evolution on the Mouse X Chromosome. Genetics. 203:1841-1857.
Larson EL, Vanderpool D, Sarver BAJ, Callahan C, Keeble S, Provencio LP, Kessler MD, Stewart V, Nordquist E, Dean MD, et al. 2018b. The Evolution of Polymorphic Hybrid Incompatibilities in House Mice. Genetics.
Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A. 103:9935-9939.
Liu C, He X, Liu W, Yang S, Wang L, Li W, Wu H, Tang S, Ni X, Wang J, et al. 2019. Bi-allelic Mutations in TTC29 Cause Male Subfertility with Asthenoteratospermia in Humans and Mice. Am J Hum Genet. 105:1168-1181.
Lorès P, Dacheux D, Kherraf Z-E, Nsota Mbango J-F, Coutton C, Stouvenel L, Ialy-Radio C, Amiri-Yekta A, Whitfield M, Schmitt A, et al. 2019. Mutations in TTC29, Encoding an Evolutionarily Conserved Axonemal Protein, Result in Asthenozoospermia and Male Infertility. Am J Hum Genet. 105:1148-1167.
Lüpold S, Manier MK, Puniamoorthy N, Schoff C, Starmer WT, Luepold SHB, Belote JM, Pitnick S. 2016. How sexual selection can drive the evolution of costly sperm ornamentation. Nature. 533:535-538.
Mack KL, Campbell P, Nachman MW. 2016. Gene regulation and speciation in house mice. Genome Res.
McKee BD, Handel MA. 1993. Sex chromosomes, recombination, and chromatin conformation. Chromosoma. 102:71-80.
McLennan HJ, Lüpold S, Smissen P, Rowe KC, Breed WG. 2017. Greater sperm complexity in the Australasian old endemic rodents (Tribe: Hydromyini) is associated with increased levels of inter-male sperm competition. Reprod Fertil Dev. 29:921-930.
Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. Proc Natl Acad Sci U S A. 100:9894-9899.
Meisel RP. 2011. Towards a More Nuanced Understanding of the Relationship between Sex-Biased Gene Expression and Rates of Protein-Coding Sequence Evolution. Mol Biol Evol. 28:1893-1900.
Meisel RP, Connallon T. 2013. The faster-X effect: integrating theory and data. Trends Genet. 29:537-544.
Meisel RP, Malone JH, Clark AG. 2012. Faster-X Evolution of Gene Expression in *Drosophila*. PLoS Genet. 8:e1003013.
Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. 2016. Contrasting Frequencies and Effects of *cis*- and *trans*-Regulatory Mutations Affecting Gene Expression. Mol Biol Evol. 33:1131-1146.
Morgan AP, Didion JP, Doran AG, Holt JM, McMillan L, Keane TM, Pardo-Manuel de Villena F. 2016. Genome Report: Whole Genome Sequence of Two Wild-Derived *Mus musculus domesticus* Inbred Strains, LEWES/EiJ and ZALENDE/EiJ, with Different Diploid Numbers. G3-Genes Genom Genet.
Namekawa SH, Park PJ, Zhang L-F, Shima JE, McCarrey JR, Griswold MD, Lee JT. 2006. Postmeiotic Sex Chromatin in the Male Germline of Mice. Curr Biol. 16:660-667.
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 32:268-274.
Pahl T, McLennan HJ, Wang Y, Achmadi AS, Rowe KC, Aplin K, Breed WG. 2018. Sperm morphology of the Rattini - are the interspecific differences due to variation in intensity of intermale sperm competition? Reprod Fertil Dev. 30:1434-1442.
Párraga M, del Mazo J. 2000. XYbp, a novel RING-finger protein, is a component of the XY body of spermatocytes and centrosomes. Mech Dev. 90:95-101.
Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet. 14:83-87.
Phifer-Rixey M, Nachman MW. 2015. Insights into mammalian biology from the wild house mouse *Mus musculus*. Elife 4.
Piasecka B, Lichocki P, Moretti S, Bergmann S, Robinson-Rechavi M. 2013. The Hourglass and the Early Conservation Models—Co-Existing Patterns of Developmental Constraints in Vertebrates. PLoS Genet. 9:e1003476.
Pitnick S, Hosken DJ, Birkhead TR. 2009. Sperm morphological diversity. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm Biology. London: Academic Press. p. 69-149.
Pizzari T, Parker G. 2009. Sperm competition and sperm phenotype. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm Biology. London: Academic Press. p. 207-245.
Ramm SA, Schärer L. 2014. The evolutionary ecology of testicular function: size isn't everything. Biol Rev. 89:874-888.
Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-Dependent Gene Expression and Evolution of the *Drosophila* Transcriptome. Science. 300:1742-1745.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene ex-

pression data. Bioinformatics. 26.
Rohlfs RV, Nielsen R. 2015. Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. Syst Biol. 64:695-708.
Sabarís G, Laiker I, Preger-Ben Noon E, Frankel N. 2019. Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. Trends Genet. 35:423-433.
Sánchez-Ramírez S, Weiss JG, Thomas CG, Cutter AD. 2021. Widespread misregulation of inter-species hybrid transcriptomes due to sex-specific and sex-chromosome regulatory evolution. PLoS Genet. 17:e1009409.
Sarver BAJ, Keeble S, Cosart T, Tucker PK, Dean MD, Good JM. 2017. Phylogenomic Insights into Mouse Evolution Using a Pseudoreference Approach. Genome Biol Evol. 9:726-739.
Schroeder CM, Valenzuela JR, Mejia Natividad I, Hocky GM, Malik HS. 2019. A Burst of Genetic Innovation in *Drosophila* Actin-Related Proteins for Testis-Specific Function. Mol Biol Evol. 37:757-772.
Schumacher J, Herlyn H. 2018. Correlates of evolutionary rates in the murine sperm proteome. BMC Evol Biol. 18:35.
Schumacher J, Rosenkranz D, Herlyn H. 2014. Mating systems and protein–protein interactions determine evolutionary rates of primate sperm proteins. Proc R Soc Biol Sci Ser B. 281.
Sin HS, Ichijima Y, Koh E, Namiki M, Namekawa SH. 2012. Human postmeiotic sex chromatin and its impact on sex chromosome evolution. Genome Res. 22:827-836.
Skinner BM, Johnson EEP, Bacon J, Affara NA, Rathje CC, Yousafzai G, Ellis PJI, Larson EL, Kopania EEK, Good JM. 2019. A high-throughput method for unbiased quantitation and categorisation of nuclear morphology.
Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, Barthès P, Kokkinaki M, Nef S, Gnirke A, et al. 2013. Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis. Cell Reports. 3:2179-2190.
Streett DA, Petersen KR, Gerritsen AT, Hunter SS, Settles ML. 2015. expHTS: analysis of high throughput sequence data in an experimental framework. Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics; Atlanta, Georgia: Association for Computing Machinery. p. 523–524.
Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive Adaptive Evolution in Mammalian Fertilization Proteins. Mol Biol Evol. 20:18-20.
Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet. 3:137-144.
The UniProt Consortium. 2020. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D480-D489.
Thompson A, May MR, Moore BR, Kopp A. 2020. A hierarchical Bayesian mixture model for inferring the expression state of genes in transcriptomes. Proc Natl Acad Sci U S A. 117:19339-19346.
Thybert D, Roller M, Navarro FCP, Fiddes I, Streeter I, Feig C, Martin-Galvez D, Kolmogorov M, Janoušek V, Akanni W, et al. 2018. Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. Genome Res.
Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian Sperm Proteins Are Rapidly Evolving: Evidence of Positive Selection in Functionally Diverse Genes. Mol Biol Evol. 19:1973-1980.
Turner JMA. 2015. Meiotic Silencing in Mammals. Annu Rev Genet. 49:395-412.
Turner LM, Chuong EB, Hoekstra HE. 2008. Comparative Analysis of Testis Protein Evolution in Rodents. Genetics. 179:2075-2089.
Vicens A, Borziak K, Karr TL, Roldan ERS, Dorus S. 2017. Comparative Sperm Proteomics in Mouse Species with Divergent Mating Systems. Mol Biol Evol. 34:1403-1416.
Vicoso B, Charlesworth B. 2009. Effective population size and the faster-x effect: An extended model. Evolution. 63:2413-2426.
Voolstra C, Tautz D, Farbrother P, Eichinger L, Harr B. 2007. Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. Genome Res. 17:42-49.
Walters JR, Harrison RG. 2011. Decoupling of rapid and adaptive evolution among seminal fluid proteins in heliconius butterflies with divergent mating systems. Evolution. 65:2855-2871.
White-Cooper H, Doggett K, Ellis RE. 2009. The evolution of spermatogenesis. In: Birkhead TR, Hosken DJ, Pitnick S, editors. Sperm Biology. London: Academic Press. p. 151-183.
Whittington E, Forsythe D, Borziak K, Karr TL, Walters JR, Dorus S. 2017. Contrasting patterns of evolutionary constraint and novelty revealed by comparative sperm proteomic analysis in Lepidoptera. BMC Genomics. 18:931.
Whittle CA, Kulkarni A, Extavour CG. 2020. Absence of a Faster-X Effect in Beetles (Tribolium, Coleoptera). G3-Genes Genom Genet. 10:1125-1136.
Winter EE, Goodstadt L, Ponting CP. 2004. Elevated Rates of Protein Secretion, Evolution, and Disease Among Tissue-Specific Genes. Genome Res. 14:54-61.
Wistuba J, Schrod A, Greve B, Hodges JK, Aslam H, Weinbauer GF, Luetjens CM. 2003. Organization of Seminiferous Epithelium in Primates: Relationship to Spermatogenic Efficiency, Phylogeny, and Mating System. Biol Reprod. 69:582-591.
Wray GA. 2007. The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet. 8:206-216.
Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 20:1377-1419.
Wyckoff GJ, Wang W, Wu C-I. 2000. Rapid evolution of male reproductive genes in the descent of man. Nature. 403:304.
Yamaguchi R, Fujihara Y, Ikawa M, Okabe M. 2012. Mice expressing aberrant sperm-specific protein PMIS2 produce normal-looking but fertilization-incompetent spermatozoa. Mol Biol Cell. 23:2671-2679.
Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 24:1586-1591.
Yapar E, Saglican E, Dönertaş HM, Özkurt E, Yan Z, Hu H, Guo S, Erdem B, Rohlfs RV, Khaitovich P, et al. 2021. unpublished data. https://wwwbiorxivorg/content/101101/010553v2. last accessed July 12, 2021.
Zhao L, Saelao P, Jones CD, Begun DJ. 2014. Origin and Spread of de Novo Genes in *Drosophila melanogaster* Populations. Science. 343:769-772.