Unraveling patterns of disrupted gene expression across a complex tissue

Kelsie E. Hunnicutt*, Jeffrey M. Good†, and Erica L. Larson*

ORCID ID: 0000-0002-9674-0630 (KEH)
ORCID ID: 0000-0003-0707-5374 (JMG)
ORCID ID: 0000-0003-3006-645X (ELL)

* University of Denver, Department of Biological Sciences, Denver, CO, 80208
† University of Montana, Division of Biological Sciences, Missoula, MO, 59812

Running title: Gene expression in complex tissues

Corresponding Authors:
Kelsie Hunnicutt
Department of Biological Sciences
University of Denver
Denver, CO 80208
Email: Kelsie.Hunnicutt@du.edu

Erica Larson
Department of Biological Sciences
University of Denver
Denver, CO 80208
Email: Erica.Larson@du.edu

ABSTRACT

Whole tissue RNASeq is the standard approach for studying gene expression divergence in evolutionary biology and provides a snapshot of the comprehensive transcriptome for a given tissue. However, whole tissues consist of diverse cell types differing in expression profiles, and the cellular composition of these tissues can evolve across species. Here, we investigate the effects of different cellular composition on whole tissue expression profiles. We compared gene expression from whole testes and enriched spermatogenesis populations in two species of house mice, *Mus musculus musculus* and *M. m. domesticus*, and their sterile and fertile F1 hybrids, which differ in both cellular composition and regulatory dynamics. We found that cellular composition differences skewed expression profiles and differential gene expression in whole testes samples. Importantly, both approaches were able to detect large-scale patterns such as disrupted X chromosome expression although whole testes sampling resulted in decreased power to detect differentially expressed genes. We encourage researchers to account for histology in RNASeq and consider methods that reduce sample complexity whenever feasible. Ultimately, we show that differences in cellular composition between tissues can modify expression profiles, potentially altering inferred gene ontological processes, insights into gene network evolution, and processes governing gene expression evolution.

Key words: gene expression evolution, hybrid sterility, speciation, RNASeq, fluorescence-activated cell sorting

INTRODUCTION

A single genome acts as the blueprint for all of the diverse cell types that comprise a eukaryotic organism. This diversity of cellular function is achieved through the expression of individual genes orchestrated by large, layered regulatory networks (Davidson and Erwin 2006; Wittkopp 2007). Often it is through gene expression that changes to the genome are connected to higher level organismal phenotypes of primary interest, and the evolution of gene expression itself can profoundly influence a species' evolutionary trajectory (King and Wilson 1975; Carroll 2008; Stern and Orgogozo 2008). Gene expression is not a static biochemical phenotype – it is an amalgamation of expression profiles of individual cell types as genes are turned on and off across organismal space and developmental time. Bulk RNASeq of whole tissues allows us to investigate these dynamics in non-model systems with minimal genomic resources and is affordable and tractable for field-based studies (Alvarez et al. 2015). However, evolutionarily important phenotypes often manifest in complex heterogenous tissues, such as sterility in reproductive organs (Turner *et al.* 2012; Suzuki and Nachman 2015), behavioral changes in neurological tissue (Sato *et al.* 2020), or color patterning across the body (Manceau *et al.* 2011; Poelstra *et al.* 2014). Standard bulk sequencing approaches necessarily collapse the complexity inherent to gene expression in these tissues and implicitly assume equivalent proportions of cell types across different comparisons. But if the relative abundance of cell types differs between contrasts, then we may be unable to distinguish regulatory divergence from differences in cellular composition (Good et al. 2010). What are the consequences of using a whole tissue approach on expression profiles and how does this impact inferences on evolutionary divergence?

Testes are emblematic of a complex tissue and are central to reproductive divergence and speciation. Testes genes are among the most rapidly evolving at the level of protein sequence (Torgerson et al. 2002; Good and Nachman 2005; Turner et al. 2008; Larson et al. 2016) and gene expression (Brawand *et al.* 2011). Sperm, which are produced by the testes, are among the most morphologically diverse animal cells (Pitnick et al. 2009) and are critical in both prezygotic (*e.g.,* sperm competition) and postzygotic (*e.g.,* hybrid sterility) reproductive barriers between species. Studies of whole testes expression have yielded great insights into the evolution of male reproductive traits (*e.g.,* Catron and Noor 2008; Davis et al. 2015; Mack et al. 2016; Ma et al. 2018; Rafati et al. 2018), but relatively few studies have accounted for the cellular complexity of

testes, a factor which we expect to complicate evolutionary inference from whole tissues (Good et al. 2010). Testes are dominated by various stages of developing sperm, primarily postmeiotic cells (~ 70% in house mice; Bellvé et al. 1977), but also present are mitotic precursors, endothelial cells, support cells (White-Cooper et al. 2009), and even multiple types of sperm in some organisms (Whittington *et al.* 2019). The relative proportion of testes cell types is evolvable and plastic (Ramm and Schärer 2014; Ramm et al. 2014) and can vary across species (Lara *et al.* 2018), mating strategies (Firman *et al.* 2015), age (Ernst *et al.* 2019; Widmayer *et al.* 2020), and social conditions (Snyder 1967). For all of these reasons, we might expect the cellular composition of testes to differ – sometimes dramatically – between different species, populations, or experimental contrasts.

The cellular complexity of tissues is often due to the developmental complexity of the phenotypes those tissues produce. In testes, undifferentiated germ cells (spermatogonia) undergo multiple rounds of mitosis then enter meiosis (spermatocytes) where they undergo two rounds of cell division to produce four haploid cells (round spermatids). These cells then undergo dramatic postmeiotic differentiation to produce mature spermatozoa. Each of these stages has a unique gene expression profile (Shima et al. 2004; Green et al. 2018) and is subject to different selective pressures (Larson et al. 2018). Spermatogenesis in many animals has an additional layer of developmental complexity in the form of the intricate regulation of the sex chromosomes. During early meiosis in mice, the X chromosome is completely transcriptionally inactivated (meiotic sex chromosome inactivation or MSCI; Handel 2004) and remains repressed for the remainder of spermatogenesis (postmeiotic sex chromosome repression or PSCR; Namekawa et al. 2006). Bulk whole testes sequencing aggregates these diverse developmental stages, limiting our resolution into how the molecular mechanisms underlying phenotypic change are acting in a developmental context (Larson et al. 2018).

The combination of the cellular heterogeneity and developmental complexity of testes is particularly relevant in understanding the evolution of hybrid male sterility. Sterile hybrids are likely to have different testes cell composition when compared to fertile mice. For example, some sterile hybrids in crosses between house mouse subspecies have only a fourth as many postmeiotic cells (Schwahn *et al.* 2018). These differences in cell composition alone might cause what looks like differential gene expression associated with hybridization. In addition, the developmental regulation of gene expression can be disrupted in hybrids (Mack and Nachman

2017; Morgan *et al.* 2020), in particular, the disruption of MSCI (Good et al. 2010; Bhattacharyya et al. 2013; Campbell et al. 2013; Larson et al. 2017). In some mouse models, the disruption of X chromosome expression only occurs at particular stages in developing sperm (*i.e.*, in cell types where the X chromosome would normally be inactivated; Larson et al. 2017), and it is not clear how patterns of stage-specific disruption in hybrids appear in whole testes where stages exhibiting normal and disrupted X regulation are combined. Evidence for disrupted X chromosome regulation in sterile hybrids varies across taxa (Davis *et al.* 2015; Rafati *et al.* 2018), but outside of mice, most studies have been restricted to whole testes RNASeq. Although these potentially confounding factors are often acknowledged in whole tissue studies (Good et al. 2010; Turner et al. 2014; Davis et al. 2015; Mugal et al. 2020), no systematic effort has been made to distinguish how differences in cellular composition can be distinguished from underlying regulatory dynamics in hybrids using whole testes samples.

Here, we use two analogous RNASeq datasets of fertile and sterile F1 hybrids from *Mus musculus musculus* and *M. m. domesticus* house mice as a model to investigate the effects of bulk whole tissue sequencing on divergent gene expression (Mack *et al.* 2016; Larson *et al.* 2017). These subspecies form a hybrid zone in Europe where they produce subfertile hybrid males (Turner et al. 2012). F1 hybrid males from wild-derived strains differ in severity of sterility dependent on the strains and the direction of the cross (Britton-Davidian et al. 2005; Good et al. 2008; Mukaj et al. 2020), with more sterile crosses having greatly disrupted cellular composition and gene expression (Good et al. 2010; Bhattacharyya et al. 2013; Campbell et al. 2013; Turner and Harr 2014; Larson et al. 2017; Schwahn et al. 2018). As a result, we can use comparisons of fertile and sterile reciprocal F1 hybrids to disentangle the effects of differing cellular composition and disrupted regulatory processes on divergent gene expression. We first examine which cell types contribute to whole testes expression profiles then test predictions about the effects of cell type abundance on whole testes comparisons. Finally, we assess whether signatures of disrupted gene regulation during specific stages of spermatogenesis are detectable in a whole tissue approach and the consequences of whole tissue sampling on differential gene expression. Collectively, we show that inferences from comparative bulk RNASeq approaches are sensitive to changes in cellular composition in complex tissues and advocate for an increased awareness of histology and tissue morphology during study design of RNASeq in non-model systems to account for such effects.

MATERIALS AND METHODS

Mouse strains and datasets

We used gene expression data from two recently published datasets analyzing disrupted hybrid gene expression from different sample types including whole testes (SRA PRJNA286765; Mack et al. 2016) and enriched cell populations across four stages of spermatogenesis (SRA PRJNA296926; Larson et al. 2017). Both studies sequenced transcriptomes from the same wild-derived inbred strains of mice from two subspecies, *M. m. domesticus* and *M. m. musculus*, and their F1 hybrids. For each subspecies, two strains were crossed to generate intraspecific F1s to reduce the effects of inbreeding depression on fertility (Good *et al.* 2008). The *M. m. domesticus* mice were generated by crossing the strains WSB/EiJ and LEWES/EiJ (hereafter *dom*), though cross direction differed between datasets with LEWES dams for the whole testes dataset and WSB dams for the enriched cell dataset. *M. m. musculus* mice were generated by crossing the strains PWK/PhJ and CZECHII/EiJ (hereafter *mus*), with PWK dams for the whole testes dataset and CZECHII dams for the sorted cell dataset. The LEWES strain from the *M. m. domesticus* subspecies and the PWK strain from the *M. m. musculus* subspecies were reciprocally crossed to generate F1 hybrid offspring with differing severity of sterility; hybrid mice from PWK female × LEWES male crosses are mostly sterile (hereafter *sterile*), while hybrid mice from LEWES female × PWK male crosses are mostly fertile (hereafter *fertile*). Mack et al. (2016) produced RNASeq libraries from whole testes for each of the four crosses ((2 parental crosses + 2 hybrid crosses) x 3 replicates per cross, N = 12). Larson et al. (2017) used Fluorescence-Activated Cell Sorting (FACS) to isolate enriched cell populations from four different stages of spermatogenesis: Mitosis: spermatogonia (SP), Meiosis$^{\text{Before X-Inact.}}$: leptotene and zygotene spermatocytes (LZ), Meiosis$^{\text{After X-Inact.}}$: diplotene spermatocytes (DIP), and Postmeiosis: round spermatids (RS) ((2 parental crosses + 2 hybrid crosses) x 3 replicates per cross x 4 cell types per replicate, N = 48).

Read mapping and count estimation

We processed both datasets in parallel through the following pipeline. First, we used Trimmomatic v.0.38 (Bolger *et al.* 2014) to trim low quality bases from the first and last 5 bp of each read and bases averaging a Phred score of less than 15 across a 4 bp sliding window. We retained reads with a minimum length of 36 bp (Table S1). To avoid mapping bias, we aligned

trimmed reads to published pseudo-reference genomes for *M. m. musculus* and *M. m. domesticus* (Huang et al. 2007) using TopHat v.2.1.1 (Trapnell *et al.* 2009) and retained up to 250 alignments per read for multi-mapped reads (-g 250). We used Lapels v.1.1.1 to convert alignments to the reference mouse genome coordinates (build GRCm38.p6) and merged alignments with suspenders v.0.2.6 (Holt et al. 2013; Huang et al. 2014). We summarized read counts for annotated genes (Ensembl Release 96) using FeatureCounts v.1.4.4 (Liao *et al.* 2014) for reads where both reads from a pair successfully aligned to the same chromosome (-B and -C). We analyzed the count data with multi-mapped reads both excluded and included (-M) from count estimates and across all annotated genes or across protein-coding genes only. Patterns were consistent across all four approaches. All results presented used only single-mapped reads and all annotated genes unless otherwise specified.

Characterizing expression patterns

To investigate how expression differed between both datasets, we defined expressed genes as those with a minimum of one Fragment Per Kilobase of exon per Million mapped reads (FPKM) in at least 3 samples within each dataset. This restricted our gene set to 16,824 total (12,587 protein-coding) genes in the whole testes dataset and 21,762 total (14,284 protein-coding) in the sorted cell dataset. We used R v.4.0.2 for all analyses. We conducted expression analyses using the Bioconductor v.3.11 package edgeR v.3.30.3 (Robinson *et al.* 2010) and normalized the data using the scaling factor method (Anders and Huber 2010).

Effects of cellular composition on whole testes expression

To first determine which cell types were present and contributing to the expression profiles of both datasets, we tested all sample types for the expression of marker genes known to be specifically expressed in certain cell types. We selected three marker genes from seven testes cell types: spermatogonia, spermatocytes, round spermatids, elongating spermatids, Sertoli cells, epithelial cells, and Leydig cells (Raymond *et al.* 2000; Nguyen *et al.* 2002; Maekawa *et al.* 2004; Li *et al.* 2007; Green *et al.* 2018). This allowed us to assess the purity of sorted cell populations by looking for the expression of non-target cell types in sorted cell populations. We were also able to identify which cell types contributed to the expression profile of whole testes.

Next, we tested the hypothesis that differential expression of stage-specific genes in whole tissues can be caused by differences in the relative abundance of cell types between comparisons—in this case *sterile* and *fertile* F1 hybrids (Good et al. 2010). We defined sets of stage-specific genes using our sorted cell populations of each subspecies (Figs S1A and B). We considered a gene to be specific to a given cell population if its median expression (normalized FPKM) was greater than two times its median expression across all other sorted cell populations (*i.e.,* an induced gene approach as in Kousathanas et al. 2014). We then compared the expression of these stage-specific genes in whole testes of *sterile* and *fertile* hybrids. We did this separately for autosomal and X-linked genes because we expected the forces driving patterns of expression to differ between the two. For autosomal genes, we expected expression to be driven largely by differences in cell composition (*e.g.,* fewer later-stage cell types in *sterile* hybrids should lead to lower expression of stage-specific genes from later stages in *sterile* compared to *fertile* whole testes). In contrast, X chromosome inactivation is disrupted in *sterile* hybrids, which should lead to higher expression of stage-specific genes from later stages in *sterile* whole testes. For autosomal genes, we used a one-sided paired Wilcoxon signed-rank test, explicitly testing if expression of stage-specific genes from more abundant cell types (Mitosis and Meiosis$^{\text{Before X-Inact.}}$) was greater in *sterile* than in *fertile* hybrid whole testes and if expression of stage-specific genes from less abundant cell types (Meiosis$^{\text{After X-Inact.}}$ and Postmeiosis) was lower in *sterile* hybrid whole testes. Because we did not know whether the effects of differing cellular compositions or misregulation of the X chromosome would be stronger for driving expression patterns of stage-specific X-linked genes in whole testes, we used two-sided Wilcoxon tests for X-linked genes. To look for additional signatures of disrupted X-linked gene expression in both sampling approaches, we also ran ANOVAs on the number of expressed X-linked genes in each cross for each sample type then conducted posthoc Tukey's tests in R.

Differential expression analysis

We conducted differential expression analysis between *sterile* and *fertile* hybrids for all five sample types in edgeR. We fit each dataset (whole testes and sorted cells separately) with negative binomial generalized linear models with Cox-Reid tagwise dispersion estimates (McCarthy *et al.* 2012) and adjusted *P*-values to a false discovery rate (FDR) of 5% (Benjamini and Hochberg 1995). We quantified the biological coefficient of variation (BCV) of parental

248 samples and hybrid samples combined and separately for each dataset. The BCV is the square
249 root of the dispersion parameter from the negative binomial model and represents variation in
250 gene expression among replicates (McCarthy *et al.* 2012).
251 We contrasted expression between *sterile* and *fertile* hybrids so that a positive log fold-
252 change (logFC) indicated over-expression in *sterile* males. For all pairwise comparisons of
253 sample types, we assessed the number of genes overlapping between both sets of differentially
254 expressed (DE) genes and the number of DE genes unique to each sample type in the
255 comparison. We also calculated whether the direction of fold change for a particular DE gene
256 switched between sample types (*e.g.,* an up-regulated DE gene in *sterile* whole testes that was a
257 down-regulated DE gene in any of the *sterile* sorted cell populations). We extended this analysis
258 comparing the direction of DE genes between sample types to parental samples, contrasting
259 expression between *mus* and *dom* parents so that a positive logFC indicated over-expression in
260 *mus* males. We tested for enrichment of specific chromosomes for DE genes between hybrids for
261 each sample type using hypergeometric tests in R (phyper) and adjusted *P*-values to an FDR of
262 5% (Benjamini and Hochberg 1995). To reduce false positives, we used only the number of
263 autosomal DE genes as the background in the hypergeometric tests because of the known over-
264 expression of the sex chromosomes in *sterile* hybrids (following Larson et al. 2016).

RESULTS

Whole testes showed unique expression patterns

Sample type, not cross, was the main driver of differences in expression profiles between samples. All sorted cell populations and whole testes samples grouped into distinct clusters (Fig 1). Within each sample type, parents formed distinct clusters and hybrids had intermediate expression. *Sterile* and *fertile* hybrids each tended to group more closely together within each sorted cell population, but hybrid crosses were intermixed for whole testes and did not form a distinct cluster.

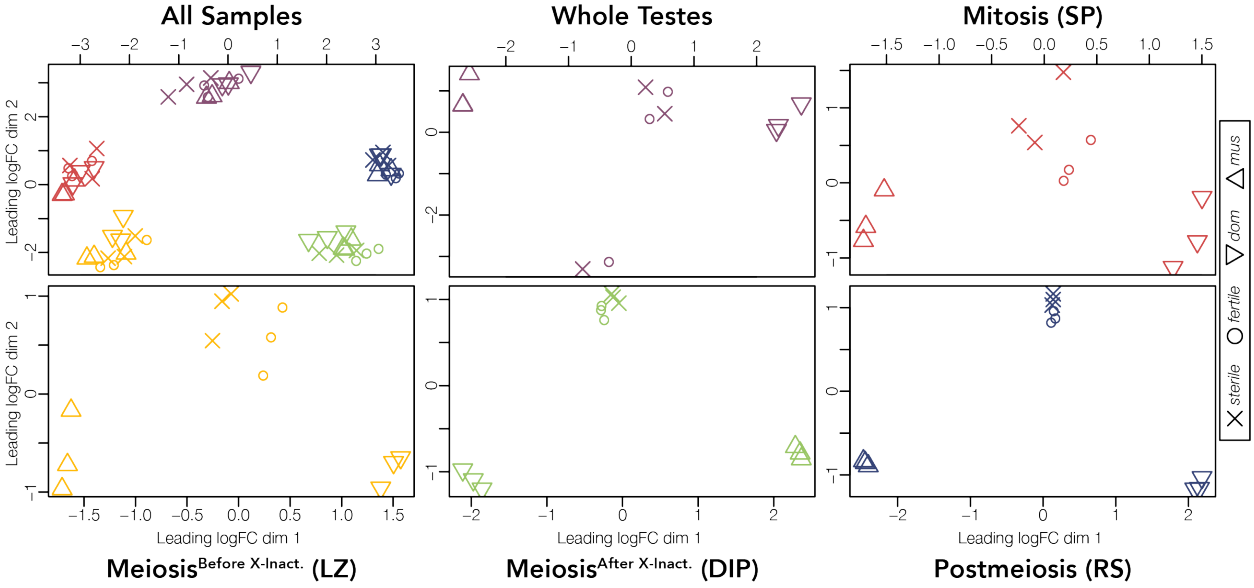

Figure 1. Sample type then cross type drives differences in expression profiles. Multidimensional scaling (MDS) plots of distances among and within sample types for expressed genes across all chromosomes. Distances are calculated as the root-mean-square deviation (Euclidean distance) of log2 fold changes among genes that distinguish each sample. Each cross is indicated by a symbol (*mus* = △, *dom* = ▽, *fertile* = O, and *sterile* = X). Samples are colored by sample type (red = Mitosis, yellow = Meiosis$^{\text{Before X-Inact.}}$, green = Meiosis$^{\text{After X-Inact.}}$, blue = Postmeiosis, and purple = Whole Testes). The upper left MDS plot includes all sample types and remaining plots show each sample type individually.

Because of the apparent increased variation among whole testes hybrid samples, we next quantified sample variation within both datasets. We measured variation among replicates using the BCV, restricting our analysis to only protein coding genes. Whole testes had greater variation

among replicates (BCV = 0.347) compared to sorted cells (BCV = 0.182). Additionally, hybrid whole testes had the greatest variation among replicates (BCV = 0.445) compared to parent whole testes (BCV = 0.207), parent sorted cells (BCV = 0.189), and hybrid sorted cells (BCV = 0.174; Fig S3). When including all annotated genes in variance calculations, the BCV was still greater in whole testes than in sorted cell populations despite the presence of some lowly expressed and highly variable non-protein coding genes in the sorted cell dataset (Figs S4-S5).

Whole testes expression patterns are driven by diverse cell composition

We next quantified expression of a panel of marker genes associated with specific testes cell types in fertile reference *mus* and *dom* samples, where gene expression is not expected to be disrupted. This allowed us to assess the purity of sorted cell populations as determined by expression of marker genes from non-target cell types and to ascertain which cell types were contributing to the unique expression patterns observed in whole testes. Our panel included marker genes associated with spermatagonia (mitosis), spermatocytes (meiosis), round spermatids (postmeiosis), elongating spermatids (postmeiosis), endothelial cells, Sertoli cells (support cells), and Leydig cells (testosterone producing cells). As expected, sorted cell populations mostly expressed only marker genes characteristic of their target cell type, overall indicating successful FACS enrichment (results for *dom* Fig 2, results for *mus* Fig S7). Mitotic cells showed high expression of spermatogonia markers and limited expression of non-target markers indicating relative cell purity. However, intermediate expression of endothelial and Sertoli markers suggested that the FACS protocol for isolating this cell population may also have captured other somatic cells. Meiotic$^{\text{Before X-Inact.}}$ cells appeared to have some spermatogonia contamination, while Meiotic$^{\text{After X-Inact.}}$ cells showed very high purity, expressing only spermatocyte-specific markers. Postmeiotic cells had high expression of round spermatid markers as expected, but also some expression of elongating spermatid markers indicating that FACS may also have captured the developmental transition to these cells.

Whole testes expressed marker genes characteristic of all seven testes cell types, especially those characteristic of postmeiotic (round and elongating spermatids) and support cell types (endothelial, Sertoli, and Leydig cells) (Fig 2). Additionally, expression patterns on the X chromosome also revealed a subset of X-linked genes unique to whole testes samples (Fig 3). This subset of genes was negligibly expressed in each of our sorted cell populations, providing

further evidence that additional cell types present in whole testes samples likely contributed to their expression profile. Mitotic (spermatogonia) and meiotic (spermatocyte) markers were also expressed in whole testes but at relatively lower FPKM values, which is consistent with the known low relative proportion of these cell types in whole testes (Bellvé *et al.* 1977; Ernst *et al.* 2019). This result suggests that early developmental cell types were contributing less to whole testes expression profiles, consistent with the hypothesis that the cellular composition of complex tissues can strongly influence relative expression levels (Good et al. 2010).

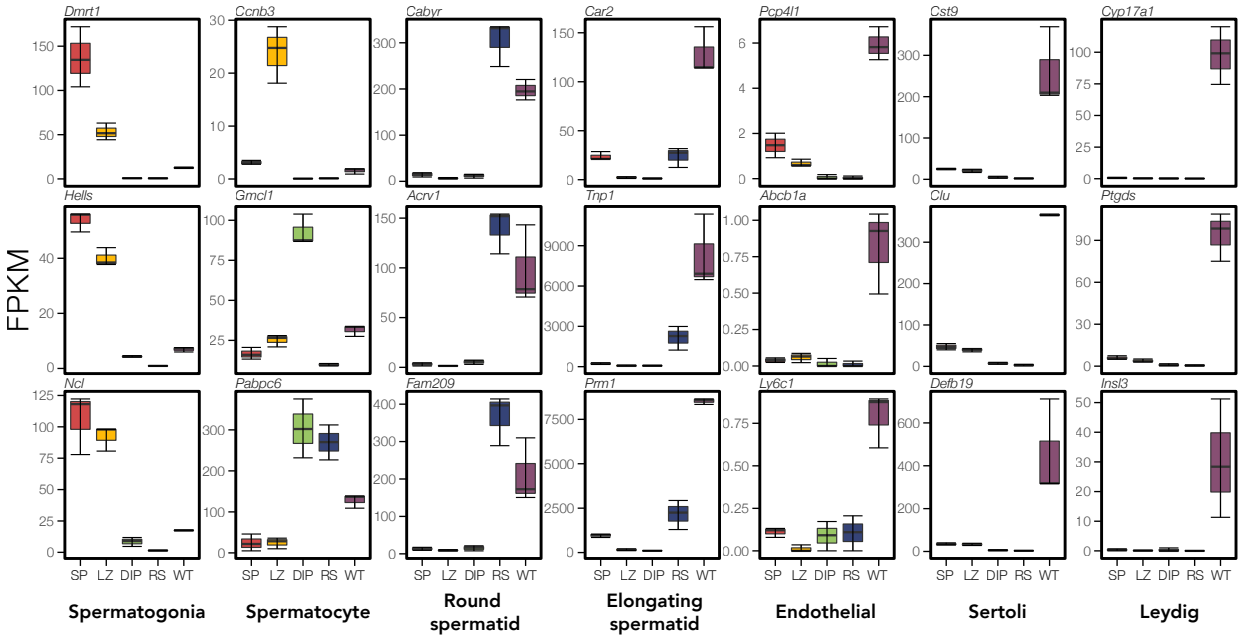

Figure 2. Whole testes expression profiles show signatures of many diverse cell types. Expression of cell-specific marker genes across each sample type for *dom* reference samples. We quantified expression (FPKM) of three marker genes (rows) associated with testes-specific cell types (columns). Each panel displays marker expression in each sample type (red = Mitosis (SP), yellow = Meiosis$^{\text{Before X-Inact.}}$ (LZ), green = Meiosis$^{\text{After X-Inact.}}$ (DIP), blue = Postmeiosis (RS), and purple = Whole Testes (WT)). Note, *Ccnb3* expression is specific to Meiotic$^{\text{Before X-Inact.}}$ cells (Maekawa et al. 2004), and *Gmc1* is specific to Meiotic$^{\text{After X-Inact.}}$ cells (Nguyen et al. 2002).

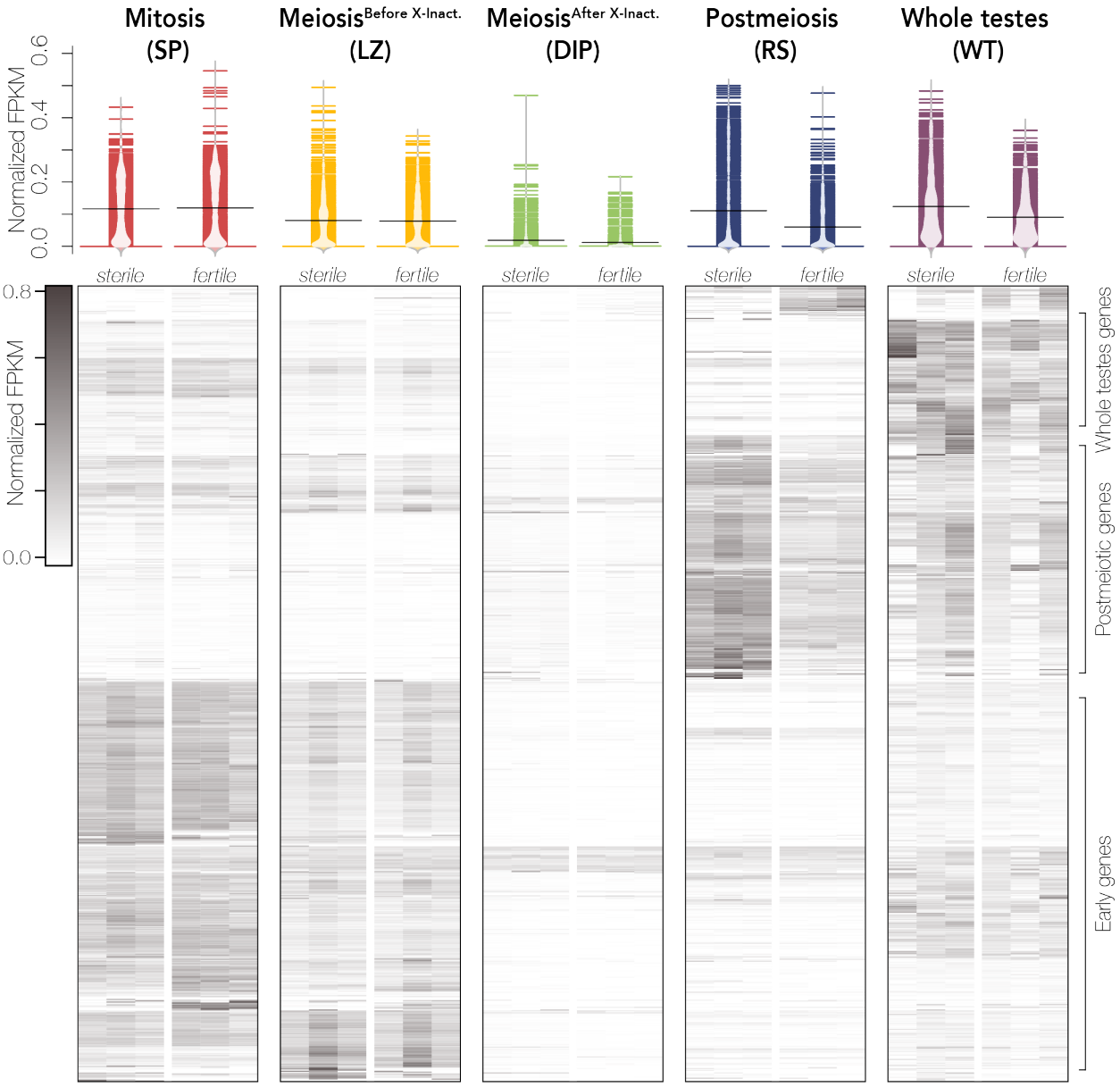

Figure 3. Patterns of X-linked gene expression in *sterile* and *fertile* hybrids differ between sorted cells and whole testes. The upper panel displays expression distributions (as normalized FPKM) across replicates for each sample type across X-linked genes. FPKM values were normalized so that the sum of squares equals one using the R package vegan (Oksanen et al. 2007). Beanplots were generated with the R package beanplot (Kampstra 2008). Beanplots are colored by sample type (red = Mitosis, yellow = Meiosis$^{\text{Before X-Inact.}}$, green = Meiosis$^{\text{After X-Inact.}}$, blue = Postmeiosis, and purple = Whole Testes) and are labelled by cross (*sterile* or *fertile* hybrid). The lower panel shows a heatmap of X-linked gene expression plotted as normalized FPKM values that are hierarchically clustered using Euclidean distance. Each row plots

expression across one gene and darker colors indicate higher expression. Heatmap was generated with the R package ComplexHeatmap v.2.3.2 (Gu *et al.* 2016).

Both changes in cellular composition of whole testes and regulatory divergence contribute to expression differences in hybrids

We further tested whether changes in cellular composition of complex tissues influences relative expression levels between contrasts. Indeed, we found that differences in whole testes cell composition between *sterile* and *fertile* hybrids appears to be a large driver of differences in relative expression of stage-specific genes (Fig 4). In *fertile* hybrids, whole testes are largely composed of late spermatogensis cell types. In *sterile* hybrids, there is a disruption in development immediately before normal MSCI, which triggers an apoptotic cascade and decreases downstream meiotic and postmeiotic cell abundance (Schwahn *et al.* 2018). Based on these histological predictions, we expected stage-specific genes from pre-X chromosome inactivation stages (Mitosis and Meiosis$^{\text{Before X-Inact.}}$) to appear over-expressed in *sterile* hybrids and stage-specific genes from post-X chromosome inactivation stages (Meiosis$^{\text{After X-Inact.}}$ and Postmeiosis) to appear under-expressed in *sterile* hybrids. Consistent with this, in whole testes, autosomal Mitotic- and Meiotic$^{\text{Before X-Inact.}}$-specific genes had higher expression in *sterile* hybrids, while autosomal Meiotic$^{\text{After X-Inact.}}$- and Postmeiotic-specific genes had lower expression (Fig 4).

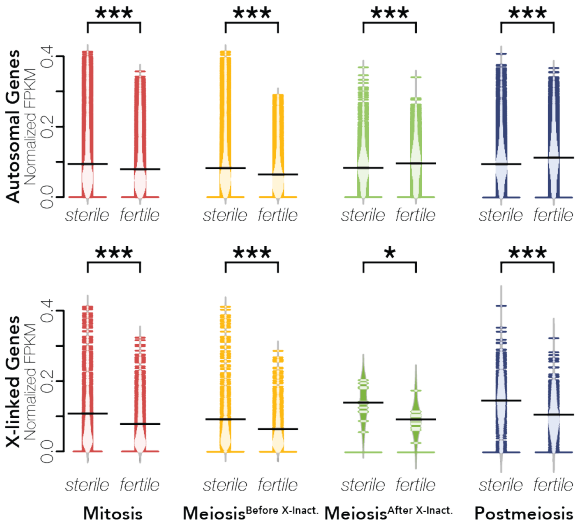

Figure 4. Changes in cellular composition alters expression of stage-specific genes in whole testes samples. For each sorted cell population, we defined a set of stage-specific genes and

compared their expression in whole testes of *sterile* and *fertile* hybrids. Mitotic and Meiotic$^{\text{Before X-Inact.}}$ cells are present at lower abundances in *sterile* hybrids while Meiotic$^{\text{After X-Inact.}}$ and Postmeiotic cells are present at higher abundances (Schwahn et al. 2018). FPKM is normalized so that the sum of squares equals 1 using the R package vegan (Oksanen et al. 2007). Differences in expression were calculated with Wilcoxon signed-rank tests where *** indicates $p < 0.001$ and * indicates $p < 0.05$ after FDR correction (Benjamini and Hochberg 1995).

Given the nature of hybrid sterility in house mice (Bhattacharyya et al. 2013), we had different expectations for X-linked genes. The normal regulation of the X chromosome is not disrupted in pre-X inactivation cell types, so differences in cellular composition should drive expression patterns for stage-specific X-linked genes in pre-X inactivation cell types as with autosomal genes. However, the X chromosome is over-expressed in post-X inactivation cell types (Larson et al. 2017), so both changes in cellular composition and known regulatory divergence could influence expression patterns of post-X inactivation stage-specific genes in *sterile* whole testes. As we predicted based on cell composition, X-linked Mitotic and Meiotic$^{\text{Before X-Inact.}}$ genes still had higher expression in *sterile* hybrids. However, X-linked Meiotic$^{\text{After X-Inact.}}$ and Postmeiotic genes also had higher expression, indicating that the disruption of X chromosome inactivation and repression in *sterile* hybrids had a stronger effect on expression patterns than changes in cell composition, despite the lower abundances of these cell types (Schwahn et al. 2018). Together these results indicate that the high proportion of postmeiotic cells in whole testes is a major cause of differences in expression patterns of autosomal and many X-linked genes between *sterile* and *fertile* whole testes samples.

We further investigated the detectability of patterns of disrupted X chromosome regulation in *sterile* hybrids across both sampling approaches and found that the whole testes sampling approach partially masks signatures of X chromosome misexpression. Previous research using sorted cell populations has shown that the inactivation of the X chromosome in *sterile* hybrids manifests as over-expression of the X chromosome both in terms of more X-linked genes being expressed as well as higher average X-linked gene expression (Larson et al. 2017). To identify if both patterns could be detected in the whole testes dataset, we compared expression of X-linked genes and calculated the number of genes expressed on the X chromosome for each cross and sample type. We recovered the expected patterns of higher

expression of X-linked genes in *sterile* hybrids in both the expected sorted cell populations (Meiosis$^{\text{After X-Inact.}}$ and Postmeiosis; Figs 2, S6A and B) and in whole testes. However, despite higher expression of X-linked genes in *sterile* whole testes, there is no significant difference in the number of expressed X-linked genes in *sterile* whole testes compared to the other cross types ($p = 0.629$; Figs 2 and S6C).

Whole testes sampling reduces power for differential expression inference

The increased variance among replicates and the resulting decreased power in the whole testes dataset also greatly reduced the number of genes considered differentially expressed between *sterile* and *fertile* hybrids in whole testes compared with sorted cell populations (Fig 5). Fewer DE genes were detected between hybrids for whole testes samples (DE genes = 83; Table S2) compared to DE genes between hybrids for each sorted cell population (Mitotic DE genes = 231, Meiotic$^{\text{Before X-Inact.}}$ DE genes = 178, Meiotic$^{\text{After X-Inact.}}$ DE genes = 343, and Postmeiotic DE genes = 606). However, both whole testes and sorted cell populations exhibited similar broad patterns of differential expression. No DE genes between *sterile* and *fertile* hybrids for each sample type were differentially up- or down-regulated in whole testes samples compared to sorted cell populations (Table S3; Fig S8). In contrast, when comparing DE genes between *mus* and *dom* mice for each sample, a small proportion of genes were differentially regulated in whole testes samples compared to sorted cell populations (0.43% - 3.16%; Table S4). In both datasets, more DE genes were upregulated in *sterile* hybrids than were downregulated (Table S3 and S4). Additionally, both datasets were able to detect enrichment of the X and Y chromosomes for DE genes as previously reported ($p < 0.001$ for both the X and Y chromosomes for all sample types; Larson et al. 2017); Fig S9; Table S2).

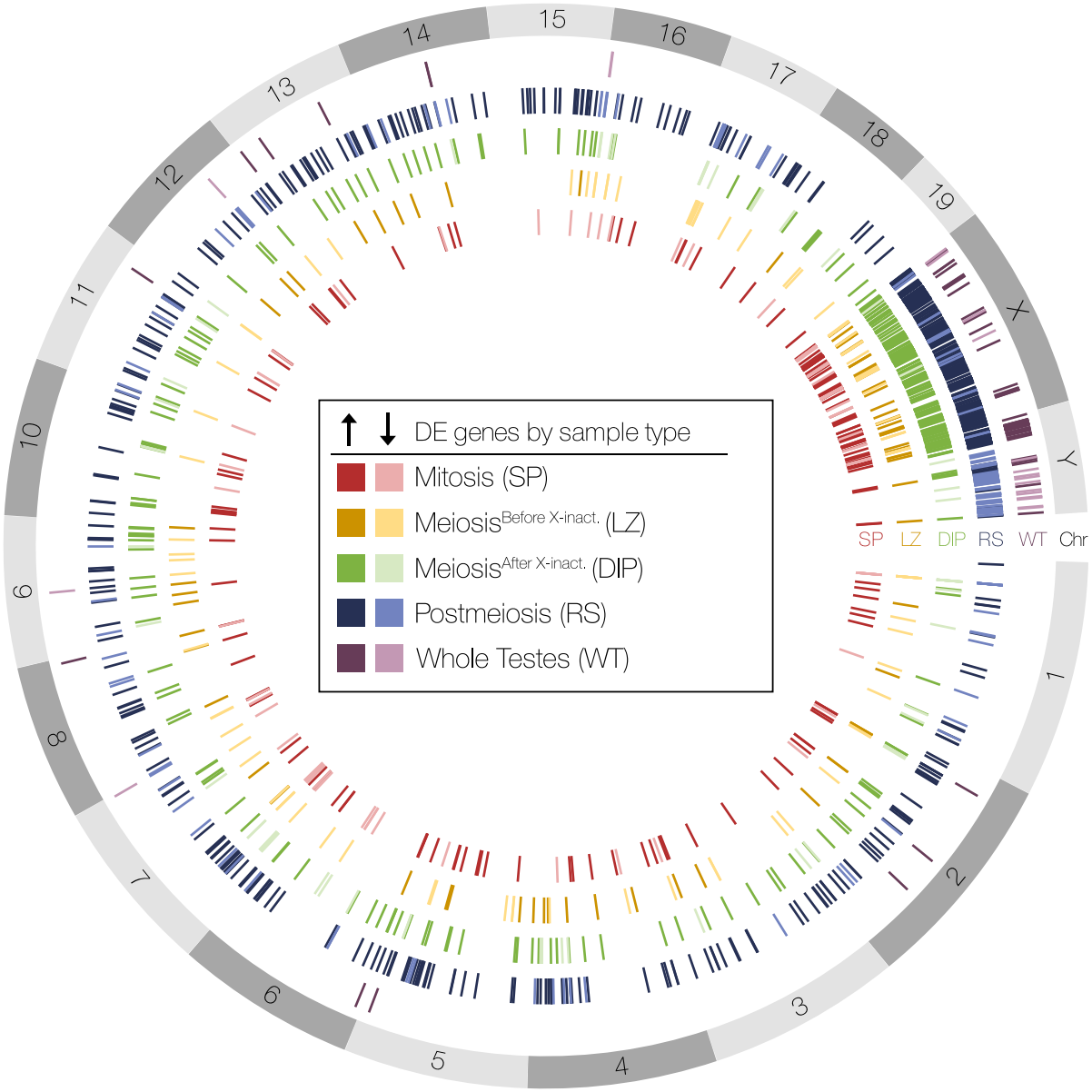

Figure 5. Whole testes and enriched cell populations differed in the number and identity of dfferentially expressed genes. Spatial distribution of differentially expressed (DE) genes across reference mouse genome chromosomes (build GRCm38.p6) between *sterile* and *fertile* hybrids for all five sample types. Darker colors indicate genes up-regulated in *sterile* hybrids and lighter colors indicated genes down-regulated in *sterile* hybrids.

Despite consistent patterns of enrichment of DE genes on the sex chromosomes and direction of expression of DE genes between hybrids, there was very little overlap in DE genes

between each sample type. Whole testes samples shared very few genes in common with any of the sorted cell populations (Fig 6). Additionally, there were very few DE genes shared across the different stages of spermatogenesis. Sorted cell samples often have large repertoires of genes that were only differentially expressed within one cell type (Fig 6) though there was greater overlap of DE genes between post-X inactivation cell types (Meiosis$^{\text{After X-Inact.}}$ and Postmeiosis). In sum, different sampling methodology clearly altered the overall and gene-specific resolution of the regulatory underpinnings of hybrid male sterility.

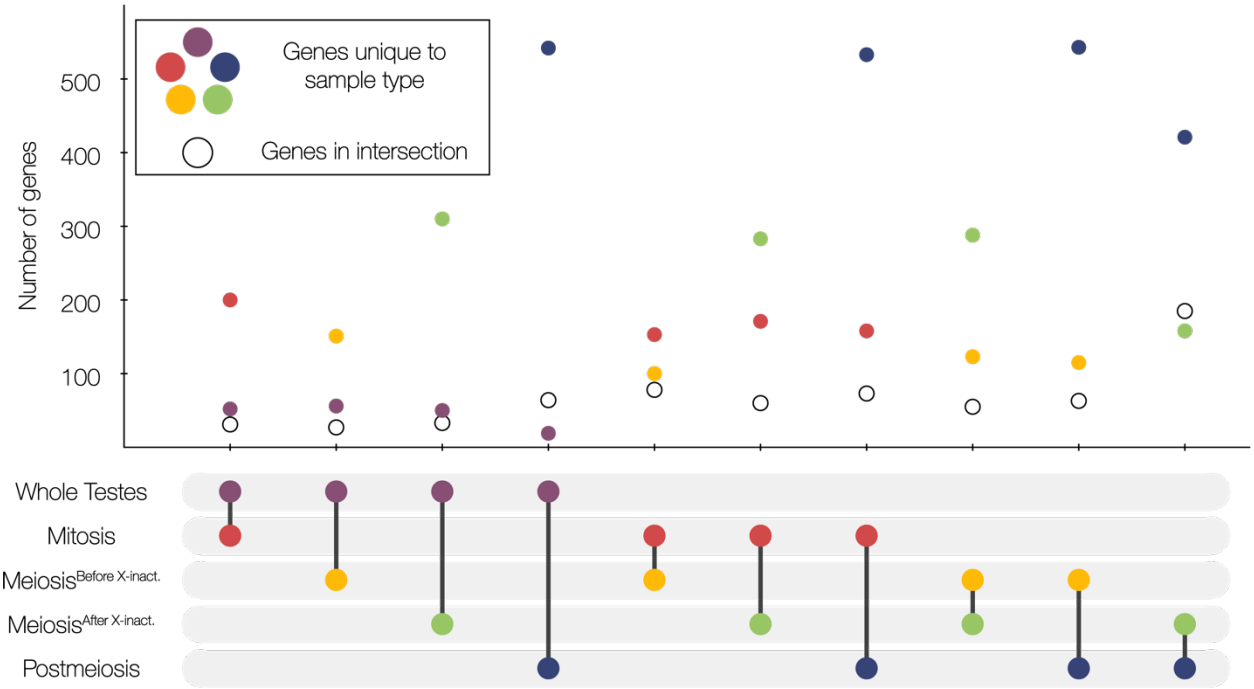

Figure 6. Whole testes and enriched cell populations differed in pairwise comparisons of DE genes between *sterile* and *fertile* hybrids. The sample types in each comparison are indicated by the pair of connected dots in the bottom panel. For each comparison, DE genes common between the two sample types are indicated with a hollow circle and DE genes unique to each sample type in that comparison are colored by sample type (red = Mitosis, yellow = Meiosis$^{\text{Before X-Inact.}}$, green = Meiosis$^{\text{After X-Inact.}}$, blue = Postmeiosis, and purple = Whole Testes).

DISCUSSION

Transcriptomic biases of complex tissues in evolutionary biology

Bulk RNASeq of whole tissues has been the canonical method for characterizing divergent expression in evolutionary biology as it is both cost-effective and tractable for wild

populations (Wang et al. 2009; Alvarez et al. 2015; Todd et al. 2016). Here we characterized patterns of expression divergence in *sterile* and *fertile* F1 hybrid house mice that differ in cellular composition using two approaches, whole testes sequencing and isolation of enriched cell populations across different stages of spermatogenesis. We demonstrated that bulk RNASeq of this complex tissue strongly reflected the cumulative contributions of diverse cell types and that the relative proportions of cell types in *sterile* and *fertile* hybrids influenced the expression of stage-specific genes. This suggests that differential expression in whole tissues can be due to either cell composition or regulatory divergence, and while these reflect fundamentally different mechanisms, they may be confounded in comparisons between species. This is a critical distinction given that researchers often interpret patterns of gene expression as reflecting per cell changes in transcript levels. This biological interpretation is implicit in models of expression evolution (Rohlfs and Nielsen 2015), which typically assume that cellular composition is stable across species of interest. We must consider the cellular context of divergent gene expression patterns (Breschi et al. 2017; Buchberger et al. 2019), as the tissues in which these phenotypes occur, such as reproductive organs (Ramm and Schärer 2014), nervous tissues (Carlson et al. 2011; Davidson and Balakrishnan 2016), and plumage (Abolins-Abols et al. 2018; Price-Waldman et al. 2020), may be prone to structural evolution, making them extremely susceptible to these confounded mechanisms inherent to whole tissue sampling.

Reproductive tissues are likely to be particularly prone to structural divergence, as cellular composition is expected to evolve in response to selection for increased reproductive success. For example, sperm competition leads to selection for males to increase sperm numbers (Firman et al. 2013, 2018). Sperm production can be increased in multiple ways, each of which has different consequences for the cellular architecture of the testis (Schärer *et al.* 2011; Ramm and Schärer 2014). The non-sperm-producing tissue within the testes can also evolve in response to sexual selection. An extreme example are Capybara, which devote ~30% of their testes to the testosterone-producing Leydig cells (in sharp contrast to other rodents, where Leydig cells comprise only 2.7-5.3% of testes; Costa et al. 2006; Lara et al. 2018). Differences in reproductive investment can also drive apparent expression differences between species. Gene expression divergence between humans and chimpanzees is elevated in testes relative to other tissues, a pattern proposed to reflect positive selection on gene expression levels (Khaitovich et al. 2005, 2006). However, whole testis transcriptomes tend to be more similar between species

with similar mating systems and cellular architectures (Brawand et al. 2011; Yapar et al. 2021), which have presumably evolved convergently in response to investment in sperm production. Our results show that even minor testis cell types (such as Leydig cells and Sertoli cells) contribute to overall expression profiles of bulk tissues and suggest that differences in the proportion of any cell type between two contrasts has the potential to strongly modify the overall expression profiles of whole tissues.

Reducing sample complexity in evolutionary studies of expression divergence

Here we confirm that FACS is an effective way of isolating relatively pure cell types and removing the effects of divergent cellular composition from experimental contrasts (Getun et al. 2011; da Cruz et al. 2016; Larson et al. 2016, 2017), although there are many other alternative methods available to assist researchers in enriching samples for cell types of interest. For example, studies of spermatogenesis initially relied on gradient centrifugation to separate out different testes cell types (Shima et al. 2004; Chalmel et al. 2007; Rolland et al. 2009), while more recent studies have increasingly relied on FACS (reviewed in Geisinger et al. 2021). Both of these approaches are well suited for studies on testes, given the dramatic changes in cell size, DNA content, and chromatin condensation that occur across spermatogenesis (Bellvé 1993; Getun et al. 2011). Various means of mechanical or flow cytometry-based isolation have been developed for enrichment of specific cell populations in other complex heterogeneous tissues (e.g., late term placenta; Li et al. 2020). Beyond these limited and potentially tissue-specific methods of bulk cell enrichment, recent advances in single cell sequencing technology (scRNA-Seq) can allow researchers to assay a greater number of cell types across many tissue types without *a priori* identification or labelling (Kiselev et al. 2019). However, these approaches may currently be less applicable for some non-model systems, especially for field-based studies, as they require access to flow cytometers and a short timeline from the time of tissue biopsy to cell sorting and RNA extraction (Getun et al. 2011; Bageritz and Raddi 2019).

When cell enrichment protocols are not feasible, alternative methods are available for minimizing developmental or cellular complexity differences between species or experimental contrasts. For example, different stages of sperm development can be isolated by sampling whole testes at different points in early sexual development (Schultz et al. 2003; Shima et al. 2004; Laiho et al. 2013), across annual reproductive cycles (Rolland et al. 2009), or spatially, as in

Drosophila, where sperm develop in tubular testes, allowing dissection of distinct regions that are enriched for particular cell types (Meiklejohn et al. 2011; Landeen et al. 2016). Furthermore, some developmentally heterogenous samples can be artificially synchronized, for example by shaving hair or plucking feathers and sampling across regrowth timelines (Poelstra et al. 2014, 2015; Ferreira et al. 2017). Microdissection of complex tissues is also a feasible way to minimize the effects of cellular composition on transcriptomic profiles. For example, laser capture microdissection provides a means to rapidly and precisely isolate cellular populations from complex tissues (Emmert-Buck et al. 1996), albeit with the added requirement of highly specialized instrumentation. It is common in behavioral research to dissect out major regions of the brain rather than sampling the whole brain (Khrameeva et al. 2020; Sato et al. 2020). Thus, a chemical or mechanical approach to partitioning complex tissues can provide researchers with a way of minimizing the negative effects associated with bulk RNASeq in their own studies.

Despite the potentially confounding effects of cellular composition and regulatory divergence in whole tissue sampling, a bulk RNASeq approach is appropriate in cases where a cell type of interest is not easily isolated or when researchers wish to capture all developmental stages. For example, Larson et al. (2017) used FACS to isolate only four stages of spermatogenesis, but postzygotic isolation barriers can operate at many different stages of spermatogenesis (Oka et al. 2010; Ishishita et al. 2015; Torgasheva and Borodin 2016; Schwahn et al. 2018; Yoshikawa et al. 2018; Liang and Sharakhov 2019). In these situations, bulk RNASeq can allow researchers to investigate expression differences in hard to obtain cell types. Additionally, some evolutionary inferences may be robust to sampling strategy. Mack et al. (2016) found that genes misexpressed in hybrid whole testes were likely disrupted because of widespread compensatory evolution, a finding that is less likely to be biased by sampling type unless the mode of evolution is expected to differ by cell type or developmental stage. Furthermore, the misexpressed genes in hybrids identified by Mack et al. (2016) overlapped substantially with sterility eQTLs identified in wild hybrids from natural hybrid zones of *M. m. musculus* and *M. m. domesticus* populations (Turner and Harr 2014), suggesting that despite the decreased power and susceptibility to artifacts introduced by differences in cellular composition associated with bulk tissue sampling, the genes that are identified are likely genes of large effect and have a high likelihood of being biologically meaningful. For all these reasons, bulk tissue

sampling may be an appropriate first step depending on the system and questions being addressed.

It is also possible to use computational approaches, such as *in silico* deconvolution methods to estimate changes in cell type proportions across samples or quantify cell type-specific expression profiles (Shen-Orr and Gaujoux 2013; Avila Cobos et al. 2018; Newman et al. 2019). These methods rely on expression profiles from single-cell data and accurate estimates of cellular proportions (Shen-Orr and Gaujoux 2013; Avila Cobos et al. 2018), which can be challenging to obtain in non-model systems but are likely to become increasingly more accessible as technologies advance. Deconvolution may also be less accurate when the expression of specific genes varies across stages because the net expression of a gene in a whole tissue may differ from its stage-specific expression. While we found that DE genes between *sterile* and *fertile* hybrids had consistent direction of differential expression between our whole testes samples and sorted cell populations, in our comparisons of DE genes between *mus* and *dom* mice, we found DE genes that had the opposite regulation patterns between sample types. Deconvolution methods in studies of hybrid misexpression may also be inherently flawed given that there is often no single "sterile" phenotype (Good et al. 2008; Turner et al. 2012; Larson et al. 2017; Bikchurina et al. 2018) and that the reference expression profiles used for deconvolution may be disrupted in hybrids (Landeen et al. 2016; Morgan et al. 2020; Mugal et al. 2020; Brekke et al. 2021). Given these drawbacks, we advocate that detailed histological analysis of how the phenotype of interest manifests in complex, heterogenous tissues (Oka et al. 2010; Schwahn et al. 2018) should accompany any evolutionary study based on comparative transcriptomic data, so that researchers can mediate biases associated with sampling methodology when designing future studies.

Power to detect differential expression using bulk RNASeq

The primary analytical goal of most RNASeq studies is to identify DE genes. It is vital that we can accurately determine which genes are differentially expressed because we use these patterns for a myriad of downstream analyses. Accurate assessment should also increase resolution into the genomic basis of phenotypes of interest. We found that bulk RNASeq can hinder differential expression analyses through an increase in replicate variability, potentially masking biologically meaningful changes in gene expression. RNASeq analyses are sensitive to

both technical and biological variation (Todd et al. 2016), and studies of outbred wild populations are inherently disadvantaged because of the power lost from increased biological variation (Liu et al. 2014; Todd et al. 2016). The BCV is an estimate of the variation among biological replicates and is correlated with power to detect DE genes. We found that in inbred strains of house mice, whole testes had higher BCV than sorted cell populations and values closer to what would be expected for an outbred wild population (BCVs greater then 0.3; Todd et al. 2016) than for genetically identical model organisms (BCV less than 0.2). Consistent with this, we found that fewer genes were differentially expressed in whole testes samples than in sorted cell populations, and those that were had little overlap with DE genes in sorted cell populations. The decreased power in bulk RNASeq is only exacerbated by the greater genetic diversity among samples from outbred wild populations, and therefore it is important to try to reduce the cellular complexity of sampled tissues.

Ultimately, both whole tissue and cell enrichment-based approaches were able to detect broad-scale patterns of disrupted sex chromosome expression in *sterile* hybrids. In house mice, MSCI is disrupted in *sterile* hybrids (Bhattacharyya et al. 2013; Davies et al. 2016; Gregorova et al. 2018), leading to an over-expression of X-linked genes (Good et al. 2010; Campbell et al. 2013; Turner et al. 2014). Both Mack *et al.* (2016) and Larson *et al.* (2017) found higher expression of genes across the X chromosome in *sterile* hybrids, but our results show that we can only detect the expression of a greater number of X-linked DE genes between *sterile* hybrids and their parents by using sorted cell populations. This pattern of over-expression can also be recovered in whole testes given *a priori* knowledge of stage-specific genes for cell types where the X chromosome should be inactivated or repressed. Of course, approaches relying on orthologous sets of stage-specific genes from other species will be limited to species with close evolutionary relationships to model organisms. A sensitivity of the regulatory mechanisms controlling sex chromosome expression during male meiosis has been proposed to be a major mechanism underlying hybrid sterility (Lifschytz and Lindsley 1972), but so far, genomic evidence for disrupted MSCI and downstream postmeiotic repression in other mammalian taxa is conflicting. In sterile hybrid cats, there is evidence of a misexpressed X chromosome (Davis *et al.* 2015), while sterile rabbit hybrids do not support a role of X chromosome misexpression in speciation (Rafati *et al.* 2018). Both studies relied on bulk whole testes sequencing and understanding if the detected or undetected misexpression of the X is biologically accurate is

important for determining the role of disrupted sex chromosome regulation in postzygotic isolation and speciation. Using targeted approaches can give us the developmental perspective needed for contextualizing the origins of reproductive barriers (Cutter and Bundus 2020).

Conclusions

Here, we demonstrate important consequences of differing cell composition in identifying DE genes in the context of hybrid sterility. We advocate for sampling approaches which allow for developmental perspectives in RNASeq studies, so that we can accurately probe species barriers. These same issues are important for other evolutionary contrasts in complex tissues, and we underscore the importance of considering the cellular and developmental context of complex expression in evolutionary studies. Our results suggest that sampling methodology could influence the biological implications of not only hybrid misexpression in speciation, but also across studies of divergent gene expression broadly. The consequences of whole tissue sampling of complex tissues have the potential to alter not only inferred gene ontological processes, but also the structure and evolution of gene networks, the relative importance of cis- and trans-regulatory evolution, and even insights into the processes and rates underlying expression evolution.

Acknowledgements: This work was supported by an NSF Graduate Research Fellowship to KEH, an NSF grant to ELL (DEB 1557059), and 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. We would like to thank Jonathan Velotta and members of the Larson and Tinghitella Labs for feedback on this project and Ben Fotovich, Ivan Kovanda, and the IT Department at the University of Denver for support using the High Performance Computing Cluster.

Author contributions: KEH, ELL, and JMG conceived of the study. KEH conducted the analyses. KEH and ELL wrote the manuscript with input from JMG.

Data accessibility: There is no data to be archived.

LITERATURE CITED

Abolins-Abols, M., E. Kornobis, P. Ribeca, K. Wakamatsu, M. P. Peterson, E. D. Ketterson, and B. Milá. 2018. Differential gene regulation underlies variation in melanic plumage coloration in the dark-eyed junco (*Junco hyemalis*). Mol. Ecol. 27:4501–4515.

Alvarez, M., A. W. Schrey, and C. L. Richards. 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? Mol. Ecol. 24:710–725.

Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. Nat. Prec., doi: 10.1038/npre.2010.4282.2.

Avila Cobos, F., J. Vandesompele, P. Mestdagh, and K. De Preter. 2018. Computational deconvolution of transcriptomics data from mixed cell populations. Bioinformatics 34:1969–1979.

Bageritz, J., and G. Raddi. 2019. Single-cell RNA sequencing with Drop-Seq. Methods Mol. Biol. 1979:73–85.

Bellvé, A. R. 1993. Purification, culture, and fractionation of spermatogenic cells. Pp. 84–113 *in* P. M. Wassarman and M. L. DePamphilis, eds. Methods in Enzymology. Academic Press, London.

Bellvé, A. R., J. C. Cavicchia, C. F. Millette, D. A. O'Brien, Y. M. Bhatnagar, and M. Dym. 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. J. Cell Biol. 74:68–85.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57:289–300.

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

Bikchurina, T. I., K. V. Tishakova, E. A. Kizilova, S. A. Romanenko, N. A. Serdyukova, A. A. Torgasheva, and P. M. Borodin. 2018. Chromosome synapsis and recombination in male-sterile and female-fertile interspecies hybrids of the dwarf hamsters (*Phodopus*, Cricetidae). Genes 9:227.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Brawand, D., M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, P. Harrigan, M. Weier, A. Liechti, A. Aximu-Petri, M. Kircher, F. W. Albert, U. Zeller, P. Khaitovich, F. Grützner, S. Bergmann, R. Nielsen, S. Pääbo, and H. Kaessmann. 2011. The evolution of gene expression levels in mammalian organs. Nature 478:343–348.

Brekke, T. D., E. C. Moore, S. C. Campbell-Staton, C. M. Callahan, Z. A. Cheviron, and J. M. Good. 2021. X chromosome-dependent disruption of placental regulatory networks in hybrid dwarf hamsters. Genetics 218:iyab043.

Breschi, A., T. R. Gingeras, and R. Guigó. 2017. Comparative transcriptomics in human and mouse. Nat. Rev. Genet. 18:425–440.

Britton-Davidian, J., F. Fel-Clair, J. Lopez, P. Alibert, and P. Boursot. 2005. Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. Biol. J. Linn. Soc. Lond. 84:379–393.

Buchberger, E., M. Reis, T.-H. Lu, and N. Posnien. 2019. Cloudy with a chance of insights: context dependent gene regulation and implications for evolutionary studies. Genes 10:492.

Campbell, P., J. M. Good, and M. W. Nachman. 2013. Meiotic sex chromosome inactivation is disrupted in sterile hybrid male house mice. Genetics 193:819–828.

Carlson, B. A., S. M. Hasan, M. Hollmann, D. B. Miller, L. J. Harmon, and M. E. Arnegard. 2011. Brain evolution triggers increased diversification of electric fishes. Science 332:583–586.

Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25–36.

Catron, D. J., and M. A. F. Noor. 2008. Gene expression disruptions of organism versus organ in *Drosophila* species hybrids. PLoS One 3:e3009.

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

Costa, D. S., T. A. R. Paula, and S. L. P. Matta. 2006. The intertubular compartment morphometry in capybaras (*Hydrochoerus hydrochaeris*) testis. Anim. Reprod. Sci. 91:173–179.

Cutter, A. D., and J. D. Bundus. 2020. Speciation and the developmental alarm clock. Elife 9:e56276.

da Cruz, I., R. Rodríguez-Casuriaga, F. F. Santiñaque, J. Farías, G. Curti, C. A. Capoano, G. A. Folle, R. Benavente, J. R. Sotelo-Silveira, and A. Geisinger. 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.

Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311:796–800.

Davidson, J. H., and C. N. Balakrishnan. 2016. Gene regulatory evolution during speciation in a songbird. G3 6:1357–1364.

Davies, B., E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli, D. Biggs, R. Diaz, C. Preece, R. Li, E. Bitoun, K. Brick, C. M. Green, R. D. Camerini-Otero, S. R. Myers, and P. Donnelly. 2016. Re-engineering the zinc fingers of *PRDM9* reverses hybrid sterility in mice. Nature 530:171–176.

Davis, B. W., C. M. Seabury, W. A. Brashear, G. Li, M. Roelke-Parker, and W. J. Murphy. 2015. Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. Mol. Biol. Evol. 32:2534–2546.

Emmert-Buck, M. R., R. F. Bonner, P. D. Smith, R. F. Chuaqui, Z. Zhuang, S. R. Goldstein, R. A. Weiss, and L. A. Liotta. 1996. Laser capture microdissection. Science 274:998–1001.

Ernst, C., N. Eling, C. P. Martinez-Jimenez, J. C. Marioni, and D. T. Odom. 2019. Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. Nat. Commun. 10:1251.

Ferreira, M. S., P. C. Alves, C. M. Callahan, J. P. Marques, L. S. Mills, J. M. Good, and J. Melo-Ferreira. 2017. The transcriptional landscape of seasonal coat colour moult in the snowshoe hare. Mol. Ecol. 26:4173–4185.

Firman, R. C., F. Garcia-Gonzalez, L. W. Simmons, and G. I. André. 2018. A competitive environment influences sperm production, but not testes tissue composition, in house mice. J. Evol. Biol. 31:1647–1654.

Firman, R. C., F. Garcia-Gonzalez, E. Thyer, S. Wheeler, Z. Yamin, M. Yuan, and L. W. Simmons. 2015. Evolutionary change in testes tissue composition among experimental populations of house mice. Evolution 69:848–855.

Firman, R. C., I. Klemme, and L. W. Simmons. 2013. Strategic adjustments in sperm production within and between two island populations of house mice. Evolution 67:3061–3070.

Geisinger, A., R. Rodríguez-Casuriaga, and R. Benavente. 2021. Transcriptomics of meiosis in the male mouse. Front Cell Dev Biol 9:446.

Getun, I. V., B. Torres, and P. R. J. Bois. 2011. Flow cytometry purification of mouse meiotic cells. J. Vis. Exp. e2602.

Good, J. M., T. Giger, M. D. Dean, and M. W. Nachman. 2010. Widespread over-expression of the X chromosome in sterile F1 hybrid mice. PLoS Genet. 6:e1001148.

Good, J. M., M. A. Handel, and M. W. Nachman. 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. Evolution 62:50–65.

Good, J. M., and M. W. Nachman. 2005. Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. Mol. Biol. Evol. 22:1044–1052.

Green, C. D., Q. Ma, G. L. Manske, A. N. Shami, X. Zheng, S. Marini, L. Moritz, C. Sultan, S. J. Gurczynski, B. B. Moore, M. D. Tallquist, J. Z. Li, and S. S. Hammoud. 2018. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-Seq. Dev. Cell 46:651-667.e10.

Gregorova, S., V. Gergelits, I. Chvatalova, T. Bhattacharyya, B. Valiskova, V. Fotopulosova, P. Jansa, D. Wiatrowska, and J. Forejt. 2018. Modulation of *Prdm9*-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. Elife 7:e34282.

Gu, Z., R. Eils, and M. Schlesner. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32:2847–2849.

Handel, M. A. 2004. The XY body: a specialized meiotic chromatin domain. Exp. Cell Res. 296:57–63.

Holt, J., S. Huang, L. McMillan, and W. Wang. 2013. Read annotation pipeline for high-throughput sequencing data. Pp. 605–612 *in* Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics - BCB'13. ACM Press, New York, New York, USA.

Huang, S., J. Holt, C.-Y. Kao, L. McMillan, and W. Wang. 2014. A novel multi-alignment pipeline for high-throughput sequencing data. Database 2014:bau057.

Huang, S., C.-Y. Kao, L. McMillan, and W. Wang. 2007. Transforming genomes using MOD files with applications. Pp. 595–604 *in* Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics - BCB'13. ACM Press, New York, New York, USA.

Ishishita, S., K. Tsuboi, N. Ohishi, K. Tsuchiya, and Y. Matsuda. 2015. Abnormal pairing of X and Y sex chromosomes during meiosis I in interspecific hybrids of *Phodopus campbelli* and *P. sungorus*. Sci. Rep. 5:9435.

Kampstra, P. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. J. Stat. Softw. 28.

Khaitovich, P., I. Hellmann, W. Enard, K. Nowick, M. Leinweber, H. Franz, G. Weiss, M. Lachmann, and S. Pääbo. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science 309:1850–1854.

Khaitovich, P., J. Kelso, H. Franz, J. Visagie, T. Giger, S. Joerchel, E. Petzold, R. E. Green, M. Lachmann, and S. Pääbo. 2006. Functionality of intergenic transcription: an evolutionary comparison. PLoS Genet. 2:e171.

Khrameeva, E., I. Kurochkin, D. Han, P. Guijarro, S. Kanton, M. Santel, Z. Qian, S. Rong, P. Mazin, M. Sabirov, M. Bulat, O. Efimova, A. Tkachev, S. Guo, C. C. Sherwood, J. G. Camp, S. Pääbo, B. Treutlein, and P. Khaitovich. 2020. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. Genome Res. 30:776–789.

King, M. C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. Science 188:107–116.

Kiselev, V. Y., T. S. Andrews, and M. Hemberg. 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. Nat. Rev. Genet. 20:273–282.

Kousathanas, A., D. L. Halligan, and P. D. Keightley. 2014. Faster-X adaptive protein evolution in house mice. Genetics 196:1131–1143.

Laiho, A., N. Kotaja, A. Gyenesei, and A. Sironen. 2013. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. PLoS One 8:e61558.

Landeen, E. L., C. A. Muirhead, L. Wright, C. D. Meiklejohn, and D. C. Presgraves. 2016. Sex chromosome-wide transcriptional suppression and compensatory cis-regulatory evolution mediate gene expression in the *Drosophila* male germline. PLoS Biol. 14:e1002499.

Lara, N. L. M., G. M. J. Costa, G. F. Avelar, S. M. S. N. Lacerda, R. A. Hess, and L. R. de França. 2018. Testis physiology—overview and histology. Pp. 105–116 *in* M. K. Skinner, ed. Encyclopedia of Reproduction. Academic Press, London.

Larson, E. L., S. Keeble, D. Vanderpool, M. D. Dean, and J. M. Good. 2017. The composite regulatory basis of the large X-effect in mouse speciation. Mol. Biol. Evol. 34:282–295.

Larson, E. L., E. E. K. Kopania, and J. M. Good. 2018. Spermatogenesis and the evolution of mammalian sex chromosomes. Trends Genet. 34:722–732.

Larson, E. L., D. Vanderpool, S. Keeble, M. Zhou, B. A. J. Sarver, A. D. Smith, M. D. Dean, and J. M. Good. 2016. Contrasting levels of molecular evolution on the mouse X chromosome. Genetics 203:1841–1857.

Li, H., Q. Huang, Y. Liu, and L. X. Garmire. 2020. Single cell transcriptome research in human placenta. Reproduction 160:R155–R167.

Li, Y.-F., W. He, K. N. Jha, K. Klotz, Y.-H. Kim, A. Mandal, S. Pulido, L. Digilio, C. J. Flickinger, and J. C. Herr. 2007. *FSCB*, a novel protein kinase A-phosphorylated calcium-binding protein, is a *CABYR*-binding partner involved in late steps of fibrous sheath biogenesis. J. Biol. Chem. 282:34104–34119.

Liang, J., and I. V. Sharakhov. 2019. Premeiotic and meiotic failures lead to hybrid male sterility in the *Anopheles gambiae* complex. Proc. Biol. Sci. 286:20191080.

Liao, Y., G. K. Smyth, and W. Shi. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

Lifschytz, E., and D. L. Lindsley. 1972. The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. U. S. A. 69:182–186.

Liu, Y., J. Zhou, and K. P. White. 2014. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30:301–304.

Ma, S., A. S. Avanesov, E. Porter, B. C. Lee, M. Mariotti, N. Zemskaya, R. Guigo, A. A. Moskalev, and V. N. Gladyshev. 2018. Comparative transcriptomics across 14 *Drosophila* species reveals signatures of longevity. Aging Cell 17:e12740.

Mack, K. L., P. Campbell, and M. W. Nachman. 2016. Gene regulation and speciation in house mice. Genome Res. 26:451–461.

Mack, K. L., and M. W. Nachman. 2017. Gene regulation and speciation. Trends Genet. 33:68–80.

Maekawa, M., C. Ito, Y. Toyama, F. Suzuki-Toyota, T. Kimura, T. Nakano, and K. Toshimori. 2004. Stage-specific expression of mouse germ cell-less-1 (mGCL-1), and multiple deformations during *mgcl-1* deficient spermatogenesis leading to reduced fertility. Arch. Histol. Cytol. 67:335–347.

Manceau, M., V. S. Domingues, R. Mallarino, and H. E. Hoekstra. 2011. The developmental role of Agouti in color pattern evolution. Science 331:1062–1065.

McCarthy, D. J., Y. Chen, and G. K. Smyth. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40:4288–4297.

Meiklejohn, C. D., E. L. Landeen, J. M. Cook, S. B. Kingan, and D. C. Presgraves. 2011. Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. PLoS Biol. 9:e1001126.

Morgan, K., B. Harr, M. A. White, B. A. Payseur, and L. M. Turner. 2020. Disrupted gene networks in subfertile hybrid house mice. Mol. Biol. Evol. 37:1547–1562.

Mugal, C. F., M. Wang, N. Backström, D. Wheatcroft, M. Ålund, M. Sémon, S. E. McFarlane, L. Dutoit, A. Qvarnström, and H. Ellegren. 2020. Tissue-specific patterns of regulatory changes underlying gene expression differences among *Ficedula* flycatchers and their naturally occurring F1 hybrids. Genome Res. 30:1727–1739.

Mukaj, A., J. Piálek, V. Fotopulosova, A. P. Morgan, L. Odenthal-Hesse, E. D. Parvanov, and J. Forejt. 2020. *Prdm9* intersubspecific interactions in hybrid male sterility of house mouse. Mol. Biol. Evol. 37:3423–3438.

Namekawa, S. H., P. J. Park, L.-F. Zhang, J. E. Shima, J. R. McCarrey, M. D. Griswold, and J. T. Lee. 2006. Postmeiotic sex chromatin in the male germline of mice. Curr. Biol. 16:660–667.

Newman, A. M., C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, and A. A. Alizadeh. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. 37:773–782.

Nguyen, T. B., K. Manova, P. Capodieci, C. Lindon, S. Bottega, X.-Y. Wang, J. Refik-Rogers, J. Pines, D. J. Wolgemuth, and A. Koff. 2002. Characterization and expression of mammalian cyclin b3, a prepachytene meiotic cyclin. J. Biol. Chem. 277:41960–41969.

Oka, A., A. Mita, Y. Takada, H. Koseki, and T. Shiroishi. 2010. Reproductive isolation in hybrid mice due to spermatogenesis defects at three meiotic stages. Genetics 186:339–351.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M. Suggests. 2007. The vegan package. Community ecology package 10:719.

Pitnick, S., D. J. Hosken, and T. R. Birkhead. 2009. Sperm morphological diversity. Pp. 69–149 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm Biology. Academic Press, London.

Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–1414.

Poelstra, J. W., N. Vijay, M. P. Höppner, and J. B. Wolf. 2015. Transcriptomics of colour patterning and colouration shifts in crows. Mol. Ecol. 24:4617–4628.

Price-Waldman, R. M., A. J. Shultz, and K. J. Burns. 2020. Speciation rates are correlated with changes in plumage color complexity in the largest family of songbirds. Evolution 74:1155–1169.

Rafati, N., J. A. Blanco-Aguiar, C. J. Rubin, S. Sayyab, S. J. Sabatino, S. Afonso, C. Feng, P. C. Alves, R. Villafuerte, N. Ferrand, L. Andersson, and M. Carneiro. 2018. A genomic map of clinal variation across the European rabbit hybrid zone. Mol. Ecol. 27:1457–1478.

Ramm, S. A., and L. Schärer. 2014. The evolutionary ecology of testicular function: size isn't everything. Biol. Rev. Camb. Philos. Soc. 89:874–888.

Ramm, S. A., L. Schärer, J. Ehmcke, and J. Wistuba. 2014. Sperm competition and the evolution of spermatogenesis. Mol. Hum. Reprod. 20:1169–1179.

Raymond, C. S., M. W. Murphy, M. G. O'Sullivan, V. J. Bardwell, and D. Zarkower. 2000. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. Genes Dev. 14:2587–2595.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Rohlfs, R. V., and R. Nielsen. 2015. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. Syst. Biol. 64:695–708.

Rolland, A. D., J.-J. Lareyre, A.-S. Goupil, J. Montfort, M.-J. Ricordel, D. Esquerré, K. Hugot, R. Houlgatte, F. Chalmel, and F. Le Gac. 2009. Expression profiling of rainbow trout testis development identifies evolutionary conserved genes involved in spermatogenesis. BMC Genomics 10:546.

Sato, D. X., N. Rafati, H. Ring, S. Younis, C. Feng, J. A. Blanco-Aguiar, C.-J. Rubin, R. Villafuerte, F. Hallböök, M. Carneiro, and L. Andersson. 2020. Brain transcriptomics of wild and domestic rabbits suggests that changes in dopamine signaling and ciliary function contributed to evolution of tameness. Genome Biol. Evol. 12:1918–1928.

Schärer, L., D. T. J. Littlewood, A. Waeschenbach, W. Yoshida, and D. B. Vizoso. 2011. Mating behavior and the evolution of sperm design. Proc. Natl. Acad. Sci. U. S. A. 108:1490–1495.

Schultz, N., F. K. Hamra, and D. L. Garbers. 2003. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. Proc. Natl. Acad. Sci. U. S. A. 100:12201–12206.

Schwahn, D. J., R. J. Wang, M. A. White, and B. A. Payseur. 2018. Genetic dissection of hybrid male sterility across stages of spermatogenesis. Genetics 210:1453–1465.

Shen-Orr, S. S., and R. Gaujoux. 2013. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. Curr. Opin. Immunol. 25:571–578.

Shima, J. E., D. J. McLean, J. R. McCarrey, and M. D. Griswold. 2004. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. Biol. Reprod. 71:319–330.

Snyder, R. L. 1967. Fertility and reproductive performance of grouped male mice. Pp. 458–472 *in* K. Benirschke, ed. Comparative aspects of reproductive failure. Springer, Berlin.

Stern, D. L., and V. Orgogozo. 2008. The loci of evolution: how predictable is genetic evolution? Evolution 62:2155–2177.

Suzuki, T. A., and M. W. Nachman. 2015. Speciation and reduced hybrid female fertility in house mice. Evolution 69:2468–2481.

Todd, E. V., M. A. Black, and N. J. Gemmell. 2016. The power and promise of RNA-seq in ecology and evolution. Mol. Ecol. 25:1224–1241.

Torgasheva, A. A., and P. M. Borodin. 2016. Cytological basis of sterility in male and female hybrids between sibling species of grey voles *Microtus arvalis* and *M. levis*. Sci. Rep. 6:36564.

Torgerson, D. G., R. J. Kulathinal, and R. S. Singh. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. Mol. Biol. Evol. 19:1973–1980.

Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Turner, L. M., E. B. Chuong, and H. E. Hoekstra. 2008. Comparative analysis of testis protein evolution in rodents. Genetics 179:2075–2089.

Turner, L. M., and B. Harr. 2014. Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. Elife 3:e02504.

Turner, L. M., D. J. Schwahn, and B. Harr. 2012. Reduced male fertility is common but highly variable in form and severity in a natural house mouse hybrid zone. Evolution 66:443–458.

Turner, L. M., M. A. White, D. Tautz, and B. A. Payseur. 2014. Genomic networks of hybrid sterility. PLoS Genet. 10:e1004162.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57–63.

White-Cooper, H., K. Doggett, and R. E. Ellis. 2009. The evolution of spermatogenesis. Pp. 151–183 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm Biology. Academic Press, London.

Whittington, E., T. L. Karr, A. J. Mongue, S. Dorus, and J. R. Walters. 2019. Evolutionary proteomics reveals distinct patterns of complexity and divergence between Lepidopteran sperm morphs. Genome Biol. Evol. 11:1838–1846.

Widmayer, S. J., M. A. Handel, and D. L. Aylor. 2020. Age and genetic background modify hybrid male sterility in house mice. Genetics 216:585–597.

Wittkopp, P. J. 2007. Variable gene expression in eukaryotes: a network perspective. J. Exp. Biol. 210:1567–1575.

Yapar, E., E. Saglican, H. M. Dönertaş, E. Özkurt, Z. Yan, H. Hu, S. Guo, B. Erdem, R. V. Rohlfs, P. Khaitovich, and M. Somel. 2021. Convergent evolution of primate testis transcriptomes reflects mating strategy. bioRxiv, doi: 10.1101/010553.

Yoshikawa, H., D. Xu, Y. Ino, T. Yoshino, T. Hayashida, J. Wang, R. Yazawa, G. Yoshizaki, and Y. Takeuchi. 2018. Hybrid sterility in fish caused by mitotic arrest of primordial germ cells. Genetics 209:507–521.