



HHS Public Access

Author manuscript

Evolution. Author manuscript; available in PMC 2023 February 01.

Published in final edited form as:

Evolution. 2022 February ; 76(2): 275–291. doi:10.1111/evo.14420.

Unraveling patterns of disrupted gene expression across a complex tissue

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

*University of Denver, Department of Biological Sciences, Denver, CO, 80208

†University of Montana, Division of Biological Sciences, Missoula, MO, 59812

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.

Keywords

gene expression evolution; hybrid sterility; speciation; RNASeq; fluorescence-activated cell sorting

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

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. Scripts used in the manuscript will be available upon publication at <https://github.com/KelsieHunnicutt>.

Conflict of interest: The authors declare no conflict of interest.

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; Montgomery and Mank 2016). 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 levels 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 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; Hermann 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 act 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. We expect sterile hybrids to have disrupted testes expression (Mack and Nachman 2017; Morgan *et al.* 2020), but sterile hybrids are also likely to have different testes cellular composition compared to fertile mice. For example, some sterile house mouse hybrids have only a fourth as many postmeiotic cells (Schwahn *et al.* 2018). These differences in cellular composition alone might cause what looks like differential gene regulation associated with hybridization. This is especially problematic when differences in cellular composition correspond to developmental timepoints where hybrid expression is disrupted, such as with the disruption of X chromosome inactivation at MSCI (Good et al. 2010; Bhattacharyya et al. 2013; Campbell et al. 2013; 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; Bredemeyer et al. 2021), 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 (Mack *et al.* 2016; Larson *et al.* 2017) as a model to investigate the effects of bulk whole tissue sequencing on divergent gene expression. 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). We use comparisons of fertile and sterile reciprocal F1 hybrids to disentangle

the effects of 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 in 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 the mouse subspecies *M. m. domesticus* and *M. m. musculus*, and their F1 hybrids. For each subspecies, two strains were crossed to generate intraspecific F1s to serve as parental controls, without 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, with LEWES dams for the whole testes dataset and WSB dams for the enriched cell dataset (hereafter *dom*). *M. m. musculus* mice were generated by crossing the strains PWK/PhJ and CZECHII/EiJ with PWK dams for the whole testes dataset and CZECHII dams for the sorted cell dataset (hereafter *mus*). F1 hybrid mice with differing severity of sterility were generated by reciprocally crossing LEWES and PWK; PWK female × LEWES male hybrids are mostly sterile (hereafter *sterile*), LEWES female × PWK male hybrids 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) × 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^{Before X-Inact.}: leptotene and zygotene spermatocytes (LZ), Meiosis^{After X-Inact.}: diplotene spermatocytes (DIP), and Postmeiosis: round spermatids (RS) ((2 parental crosses + 2 hybrid crosses) × 3 replicates per cross × 4 cell types per replicate, N = 48). In both studies, libraries were sequenced on an Illumina HiSeq 2000 (100 bp, PE).

Read mapping and count estimation

We processed both datasets in parallel. 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. We used Lapels v.1.1.1 to convert alignments to the reference mouse genome coordinates (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 read pairs that aligned to the same chromosome (-B and -C). We analyzed the count data with and without multi-mapped reads (-M) and across all annotated genes or protein-coding genes only. We found consistent results using all approaches and here present results using only uniquely mapped reads for all annotated genes, unless otherwise specified. Whole testes samples averaged ~24 million mapped read pairs per sample while sorted cell populations averaged ~8 million read pairs.

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 analysis to 16,824 genes (12,587 protein-coding) in the whole testes dataset and 21,762 genes (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) as well as marker genes from (Hermann *et al.* 2018). This approach 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 (Fig 1; Good *et al.* 2010). We defined sets of stage-specific genes using our sorted cell populations of each subspecies (Figs S1A, B; Supplemental File 1). 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 cellular 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 one-sided paired Wilcoxon signed-rank tests to test if expression of stage-specific genes from more abundant cell types (Mitosis and Meiosis^{Before X-Inact.}) was greater in *sterile* hybrid whole testes and if expression of stage-specific genes from less abundant cell types (Meiosis^{After X-Inact.} and Postmeiosis) was lower in *sterile* hybrid whole testes. Because we did not know whether the effects of differing cellular composition 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 signed-rank tests for X-linked genes.

To look for additional signatures of disrupted X-linked gene expression in both sampling approaches, we first used one-sided Wilcoxon signed-rank tests to compare expression of X-linked genes in *sterile* compared to *fertile* hybrids for each sample type where disrupted X-linked expression was expected (*i.e.*, Meiosis^{After X-Inact.}, Postmeiosis, and Whole Testes). Then, we defined sets of “detected” genes for each sample type as those expressed above an FPKM threshold within a replicate (FPKM > 0 – 4) and ran one-way ANOVAs on the number of detected X-linked genes in each cross within a sample type and conducted posthoc Tukey’s tests. Note, because there are only three replicates per sample type, these data inevitably violate distribution assumptions in both parametric and non-parametric tests, and differences among treatments should be considered largely qualitative.

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 samples and hybrid samples combined and separately for each dataset. The BCV is the square root of the dispersion parameter from the negative binomial model and represents variation in gene expression among replicates (McCarthy *et al.* 2012). We used two bootstrapping approaches to determine whether BCVs differed across datasets. First, we subsampled raw count files for 10000 genes across 100 replicates and recalculated the BCV for four groups: hybrid whole testes, parent whole testes, hybrid sorted cells, and parent sorted cells. Second, we dropped one individual per group and recalculated the BCV for every set of *n*-1 individuals. For both approaches, we estimated 99% confidence intervals for the bootstrap BCV estimates from each group and approach (reported as CI₁ and CI₂, respectively).

We contrasted expression between *sterile* and *fertile* hybrids so that a positive log fold-change (logFC) indicated over-expression in *sterile* males. For all pairwise comparisons of sample types, we assessed the number of genes overlapping between both sets of differentially expressed (DE) genes and the number of DE genes unique to each sample type in the comparison. We also calculated whether the direction of fold change for a particular DE gene switched between sample types (*e.g.*, an up-regulated DE gene in *sterile* whole testes that was a down-regulated DE gene in any of the *sterile* sorted cell populations).

We extended this analysis comparing the direction of DE genes between sample types to parental samples, contrasting expression between *mus* and *dom* parents so that a positive logFC indicated over-expression in *mus* males. We tested for enrichment of specific chromosomes for DE genes between hybrids for each sample type using hypergeometric tests in R (phyper) and adjusted *P*-values to an FDR of 5% (Benjamini and Hochberg 1995). To reduce false positives, we used only the number of autosomal DE genes as the background in the hypergeometric tests because of the known over-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 2). 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.

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; Fig S2). Additionally, hybrid whole testes had the greatest variation among replicates (BCV = 0.445; CI₁ = [0.443,0.445]; CI₂ = [0.375,0.514]) compared to parent whole testes (BCV = 0.207; CI₁ = [0.207,0.208]; CI₂ = [0.161,0.251]), parent sorted cells (BCV = 0.189; CI₁ = [0.190,0.191]; CI₂ = [0.185,0.192]), and hybrid sorted cells (BCV = 0.174; CI₁ = [0.175,0.176]; CI₂ = [0.171,0.177]; Fig S3-S5). 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 S6, S7).

Whole testes expression patterns are driven by diverse cellular composition

We next quantified expression of two panels 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 first panel included marker genes associated with spermatogonia (mitosis), spermatocytes (meiosis), round spermatids (postmeiosis), elongating spermatids (postmeiosis), endothelial cells, Sertoli cells (support cells), and Leydig cells (testosterone producing cells), while the second panel included additional cell types (from Hermann et al. 2018; Figs S8, S9). Results from both marker panels were consistent. As expected, sorted cell populations mostly expressed only marker genes characteristic of their target cell type, overall indicating successful FACS enrichment (results for *dom* Figs 3, S8, results for *mus* Figs S9, S10). 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^{Before X-Inact.} cells appeared to have some spermatogonia contamination, while Meiotic^{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, particularly postmeiotic (round and elongating spermatids) and support cell types (endothelial, Sertoli, and Leydig cells) (Fig 3). Additionally, expression patterns on the X chromosome revealed a subset of X-linked genes unique to whole testes samples (Fig 4). These genes were 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 low relative proportion of these cell types in whole testes (Bellvé *et al.* 1977; Ernst *et al.* 2019). This suggests that early developmental cell types contributed 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).

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 cellular composition between *sterile* and *fertile* hybrids appears to be a large driver of differences in relative expression of stage-specific genes (Fig 5). In *fertile* hybrids, whole testes are largely composed of late spermatogenesis 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^{Before X-Inact.}) to appear over-expressed in *sterile* hybrids and stage-specific genes from post-X chromosome inactivation stages (Meiosis^{After X-Inact.} and Postmeiosis) to appear under-expressed in *sterile* hybrids. Consistent with this, in whole testes, autosomal Mitotic- and Meiotic^{Before X-Inact.}-specific genes had higher expression in *sterile* hybrids (one-sided Wilcoxon Signed-Rank Test; autosomal Mitotic: $n = 5307$, $V = 11247685$, $p = 0$; autosomal Meiotic^{Before X-Inact.}: $n = 4215$, $V = 7988923$, $p = 0$), while autosomal Meiotic^{After X-Inact.}- and Postmeiotic-specific genes had lower expression (one-sided Wilcoxon Signed-Rank Test; autosomal Meiotic^{After X-Inact.}: $n = 4544$, $V = 2005025$, $p = 1.46 \times 10^{-276}$; autosomal Postmeiotic: $n = 7417$, $V = 4789686$, $p = 0$; Fig 5).

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 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 cellular composition, X-linked Mitotic and Meiotic^{Before X-Inact.} genes still had higher expression in *sterile* hybrids (one-sided Wilcoxon signed-rank test; X-linked Mitotic: $n = 465$, $V = 95946$, $p = 1.16 \times 10^{-47}$; X-linked Meiotic^{Before X-Inact.}: $n = 361$, $V = 60492$, $p = 1.53 \times 10^{-44}$). However, X-linked Meiotic^{After X-Inact.} and Postmeiotic genes also had higher expression (two-sided Wilcoxon Signed-Rank Test; X-linked Meiotic^{After X-Inact.}: $n = 11$, $V = 56$, $p = 0.0420$; X-linked Postmeiotic: $n = 252$, $V = 25826$, $p = 1.59 \times 10^{-17}$), indicating that the disruption of X chromosome inactivation and repression in *sterile* hybrids had a stronger effect on expression patterns than changes in cellular 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 whole testes sampling partially masks signatures of X chromosome misexpression. Previous research using sorted cell populations has shown that disruption of MSCI in *sterile* hybrids manifests as over-expression of the X chromosome both in terms of more expressed X-linked genes and higher average X-linked gene expression (Larson et al. 2017). We recovered the expected pattern of higher X-linked gene expression in *sterile* hybrids in both sorted cell populations (one-sided Wilcoxon signed-rank test; X-linked Meiosis^{After X-Inact.}: $n = 896$, $V = 273584$, $p = 5.34 \times 10^{-60}$; X-linked Postmeiosis: $n = 896$, $V = 290110$, $p = 1.18 \times 10^{-64}$; Fig 4) and in whole testes (one-sided Wilcoxon signed-rank test; $n = 896$, $V = 326947$, $p = 2.40 \times 10^{-64}$; Fig 4). We also found more detected X-linked genes in Meiosis^{After X-Inact.} ($F_{3,8} = 13.8$, $p = 1.58 \times 10^{-3}$; Fig S11A) and Postmeiosis ($F_{3,8} = 31.87$, $p = 8.47 \times 10^{-05}$; Fig S11B), but there was no difference in the number of detected X-linked genes in *sterile* whole testes ($F_{3,8} = 0.606$, $p = 0.629$; Fig S11C), regardless of the FPKM threshold used to define detected genes.

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 6; Table S2; Supplemental File 2). Fewer DE genes were detected between hybrids for whole testes samples (DE genes = 83; Table S2) compared to sorted cell populations (Mitotic DE genes = 231, Meiotic^{Before X-Inact.} DE genes = 178, Meiotic^{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. In both datasets, more

DE genes were upregulated in *sterile* hybrids than were downregulated (Table S2), and we were able to detect enrichment of the X and Y chromosomes for DE genes as previously reported (Larson et al. 2017; Fig S12; Tables S3, S4). In addition, no DE genes between *sterile* and *fertile* hybrids were differentially up- or down-regulated in whole testes samples compared to sorted cell populations (Table S5; Fig S13), although we did find this pattern when comparing DE genes between *mus* and *dom* mice—a small proportion of genes were differentially regulated in whole testes samples compared to sorted cell populations (0.43% - 3.16%; Table S6; Supplemental Files S3, S4).

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 7). Additionally, there were very few DE genes shared across the different stages of spermatogenesis, although the proportion of DE genes shared between sample types generally increased with stricter fold change cutoffs (Table S7). Sorted cell samples often have large repertoires of genes that were only differentially expressed within one cell type (Fig 7) though there was greater overlap of DE genes between post-X inactivation cell types (Meiosis^{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.

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 cell type proportions 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 cellular 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 (Rohlf 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 (Montgomery and Mank 2016; 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 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) or the proportion of sperm-producing tissue within the testes (Lüpold et al. 2009). 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 in bulk tissues even minor testis cell types (such as Leydig cells and Sertoli cells) contribute to overall expression profiles and suggest that differences in the proportion of any cell type have the potential to strongly modify 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; Geisinger et al. 2021; Kopania et al. 2021). There are of course, alternative methods available for bulk cell enrichment, such as gradient centrifugation to separate testes cell types (Shima et al. 2004; Chalmel et al. 2007; Rolland et al. 2009). These approaches are well suited for testes, given the dramatic changes in cell size, DNA content, and chromatin condensation during spermatogenesis (Bellvé 1993; Getun et al. 2011), but mechanical or flow cytometry-based enrichment has also been developed for other complex heterogeneous tissues (e.g., late term placenta; Li et al. 2020). There is the potential for FACS to bias gene expression in enriched cell populations (*i.e.*, by triggering a stress response from nozzle pressure or UV exposure; Box et al. 2020). However, cell sorting procedures appear to have a minimal effect on overall expression profiles, and altered expression is likely consistent across treatments within an experiment and can be mitigated by minimizing the time between cell sorting, RNA extraction, and storage (Box et al. 2020). Beyond the limited and potentially tissue-specific methods of bulk cell enrichment, recent advances in single cell sequencing technology (scRNA-Seq) and spatial transcriptomic methods (*e.g.*, sci-SPACE, Srivatsan et al. 2021) 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). Although both FACS and scRNA-Seq are both powerful approaches for studying gene expression evolution in tissues with cellular composition differences (Kopania et al. 2021; Murat et al. 2021), they are both currently difficult to apply in non-model systems, especially for field-based studies, as they typically require access to flow cytometers and a short timeline for tissue biopsy, cell sorting, and

RNA extraction (Getun et al. 2011; Bageritz and Raddi 2019; but see also Wohnhaas et al. 2019; Denisenko et al. 2020). Additionally, scRNA-Seq protocols likely have some of the same sources of biased gene expression as FACS, and further research should be done to determine how different enrichment protocols alter expression inferences.

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 heterogeneous 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. Indeed, some evolutionary inferences may be robust to sampling strategy. 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 (Landeem 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 inter-replicate variability in expression than sorted cell populations and levels of variation closer to what would be expected for an outbred wild population (BCVs greater than 0.3; McCarthy et al. 2012; Todd et al. 2016) than for genetically identical model organisms (BCV less than 0.2). We suggest that reporting BCV should become a best-practices standard for all RNASeq studies so that researchers may better understand the nature of biological variation in gene expression across a variety of evolutionary contrasts. Consistent with the increased BCV in whole testes samples, we found that fewer genes were differentially expressed in whole testes samples than in sorted cell populations and that DE genes in whole testes had little overlap with DE genes in sorted cell populations. However, this overlap proportionally increased with stricter fold change cutoffs, which strongly supports using these cutoffs to decrease the chance of detecting false positive DE genes (as proposed by Montgomery and Mank 2016). The downside to this more conservative approach was that the higher fold change cutoffs likely led to the exclusion of some genes with biologically relevant expression differences.

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 it is more difficult to detect an increased number of expressed X-linked DE genes between *sterile* hybrids and their parents using whole testes sampling. Patterns of X-linked 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, the X chromosome is misexpressed and MSCI is disrupted (Davis et al. 2015; Bredemeyer et al. 2021), while sterile rabbit hybrids do not support a role of X chromosome misexpression in speciation (Rafati *et al.* 2018). Studies outside of house mice have largely relied on bulk whole testes sequencing (but see Bredemeyer et al. 2021) 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 cellular 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work was supported by an NSF Graduate Research Fellowship to KEH (DGE- 2034612), an NSF grant to ELL (DEB 2012041), 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, Catherine Durso for statistical advice, and Ben Fotovich, Ivan Kovanda, and the IT Department at the University of Denver for support using the High Performance Computing cluster.

LITERATURE CITED

- Abolins-Abols M, Kornobis E, Ribeca P, Wakamatsu K, Peterson MP, Ketterson ED, and Milá B. 2018. Differential gene regulation underlies variation in melanic plumage coloration in the dark-eyed junco (*Junco hyemalis*). *Mol. Ecol.* 27:4501–4515. [PubMed: 30252177]
- Alvarez M, Schrey AW, and Richards CL. 2015. Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Mol. Ecol.* 24:710–725. [PubMed: 25604587]
- Anders S, and Huber W. 2010. Differential expression analysis for sequence count data. *Nat. Prec.*, doi: 10.1038/npre.2010.4282.2.
- Avila Cobos F, Vandesompele J, Mestdagh P, and De Preter K. 2018. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* 34:1969–1979. [PubMed: 29351586]
- Bageritz J, and Raddi G. 2019. Single-cell RNA sequencing with Drop-Seq. *Methods Mol. Biol.* 1979:73–85.
- Bellvé AR 1993. Purification, culture, and fractionation of spermatogenic cells. Pp. 84–113 in Wassarman PM and DePamphilis ML, eds. *Methods in Enzymology*. Academic Press, London.
- Bellvé AR, Cavicchia JC, Millette CF, O'Brien DA, Bhatnagar YM, and Dym M. 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J. Cell Biol.* 74:68–85. [PubMed: 874003]
- Benjamini Y, and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57:289–300.
- Bhattacharyya T, Gregorova S, Mihola O, Anger M, Sebestova J, Denny P, Simecek P, and Forejt J. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. *Proc. Natl. Acad. Sci. U. S. A.* 110:E468–E477. [PubMed: 23329330]
- Bikhurina TI, Tishakova KV, Kizilova EA, Romanenko SA, Serdyukova NA, Torgasheva AA, and Borodin PM. 2018. Chromosome synapsis and recombination in male-sterile and female-fertile interspecies hybrids of the dwarf hamsters (*Phodopus, Cricetidae*). *Genes* 9:227.
- Bolger AM, Lohse M, and Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. [PubMed: 24695404]
- Box A, DeLay M, Tighe S, Chittur SV, Bergeron A, Cochran M, Lopez P, Meyer EM, Saluk A, Thornton S, and Brundage K. 2020. Evaluating the Effects of Cell Sorting on Gene Expression. *J. Biomol. Tech.* 31:100–111. [PubMed: 32982601]
- Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, Albert FW, Zeller U, Khaitovich P, Grützner F, Bergmann S, Nielsen R, Pääbo S, and Kaessmann H. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348. [PubMed: 22012392]
- Bredemeyer KR, Seabury CM, Stickney MJ, McCarrey JR, vonHoldt BM, and Murphy WJ. 2021. Rapid macrosatellite evolution promotes X-Linked hybrid male sterility in a feline interspecies cross. *Mol. Biol. Evol.*, doi: 10.1093/molbev/msab274.
- Brekke TD, Moore EC, Campbell-Staton SC, Callahan CM, Cheviron ZA, and Good JM. 2021. X chromosome-dependent disruption of placental regulatory networks in hybrid dwarf hamsters. *Genetics* 218:iyab043.
- Breschi A, Gingeras TR, and Guigó R. 2017. Comparative transcriptomics in human and mouse. *Nat. Rev. Genet.* 18:425–440. [PubMed: 28479595]

- Britton-Davidian J, Fel-Clair F, Lopez J, Alibert P, and Boursot P. 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, Reis M, Lu T-H, and Posnien N. 2019. Cloudy with a chance of insights: context dependent gene regulation and implications for evolutionary studies. *Genes* 10:492.
- Campbell P, Good JM, and Nachman MW. 2013. Meiotic sex chromosome inactivation is disrupted in sterile hybrid male house mice. *Genetics* 193:819–828. [PubMed: 23307891]
- Carlson BA, Hasan SM, Hollmann M, Miller DB, Harmon LJ, and Arnegard ME. 2011. Brain evolution triggers increased diversification of electric fishes. *Science* 332:583–586. [PubMed: 21527711]
- Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134:25–36. [PubMed: 18614008]
- Catron DJ, and Noor MAF. 2008. Gene expression disruptions of organism versus organ in *Drosophila* species hybrids. *PLoS One* 3:e3009.
- Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, Gattiker A, Moore J, Patard J-J, Wolgemuth DJ, Jégou B, and Primig M. 2007. The conserved transcriptome in human and rodent male gametogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 104:8346–8351. [PubMed: 17483452]
- Costa DS, Paula TAR, and Matta SLP. 2006. The intertubular compartment morphometry in capybaras (*Hydrochoerus hydrochaeris*) testis. *Anim. Reprod. Sci.* 91:173–179. [PubMed: 16310104]
- Cutter AD, and Bundus JD. 2020. Speciation and the developmental alarm clock. *Elife* 9:e56276.
- da Cruz I, Rodríguez-Casuriaga R, Santiñaque FF, Fariñas J, Curti G, Capoano CA, Folle GA, Benavente R, Sotelo-Silveira JR, and 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. [PubMed: 27094866]
- Davidson EH, and Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. *Science* 311:796–800. [PubMed: 16469913]
- Davidson JH, and Balakrishnan CN. 2016. Gene regulatory evolution during speciation in a songbird. *G3* 6:1357–1364. [PubMed: 26976438]
- Davies B, Hatton E, Altemose N, Hussin JG, Pratto F, Zhang G, Hinch AG, Moralli D, Biggs D, Diaz R, Preece C, Li R, Bitoun E, Brick K, Green CM, Camerini-Otero RD, Myers SR, and Donnelly P. 2016. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature* 530:171–176. [PubMed: 26840484]
- Davis BW, Seabury CM, Brashear WA, Li G, Roelke-Parker M, and Murphy WJ. 2015. Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. *Mol. Biol. Evol.* 32:2534–2546. [PubMed: 26006188]
- Denisenko E, Guo BB, Jones M, Hou R, de Kock L, Lassmann T, Poppe D, Clément O, Simmons RK, Lister R, and Forrest ARR. 2020. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* 21:130. [PubMed: 32487174]
- Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, Weiss RA, and Liotta LA. 1996. Laser capture microdissection. *Science* 274:998–1001. [PubMed: 8875945]
- Ernst C, Eling N, Martinez-Jimenez CP, Marioni JC, and Odom DT. 2019. Staged developmental mapping and X chromosome transcriptional dynamics during mouse spermatogenesis. *Nat. Commun.* 10:1251. [PubMed: 30890697]
- Ferreira MS, Alves PC, Callahan CM, Marques JP, Mills LS, Good JM, and Melo-Ferreira J. 2017. The transcriptional landscape of seasonal coat colour moult in the snowshoe hare. *Mol. Ecol.* 26:4173–4185. [PubMed: 28500774]
- Firman RC, Garcia-Gonzalez F, Simmons LW, and André GI. 2018. A competitive environment influences sperm production, but not testes tissue composition, in house mice. *J. Evol. Biol.* 31:1647–1654. [PubMed: 30074655]
- Firman RC, Garcia-Gonzalez F, Thyer E, Wheeler S, Yamin Z, Yuan M, and Simmons LW. 2015. Evolutionary change in testes tissue composition among experimental populations of house mice. *Evolution* 69:848–855. [PubMed: 25600637]

- Firman RC, Klemme I, and Simmons LW. 2013. Strategic adjustments in sperm production within and between two island populations of house mice. *Evolution* 67:3061–3070. [PubMed: 24094355]
- Geisinger A, Rodríguez-Casuriaga R, and Benavente R. 2021. Transcriptomics of meiosis in the male mouse. *Front Cell Dev Biol* 9:446.
- Getun IV, Torres B, and Bois PRJ. 2011. Flow cytometry purification of mouse meiotic cells. *J. Vis. Exp.* e2602.
- Good JM, Giger T, Dean MD, and Nachman MW. 2010. Widespread over-expression of the X chromosome in sterile F1 hybrid mice. *PLoS Genet.* 6:e1001148.
- Good JM, Handel MA, and Nachman MW. 2008. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* 62:50–65. [PubMed: 18005156]
- Good JM, and Nachman MW. 2005. Rates of protein evolution are positively correlated with developmental timing of expression during mouse spermatogenesis. *Mol. Biol. Evol.* 22:1044–1052. [PubMed: 15647515]
- Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, Moritz L, Sultan C, Gurczynski SJ, Moore BB, Tallquist MD, Li JZ, and Hammoud SS. 2018. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-Seq. *Dev. Cell* 46:651–667.e10.
- Gregorova S, Gergelits V, Chvatalova I, Bhattacharyya T, Valiskova B, Fotopulosova V, Jansa P, Wiatrowska D, and Forejt J. 2018. Modulation of Prdm9-controlled meiotic chromosome asynapsis overrides hybrid sterility in mice. *Elife* 7:e34282.
- Gu Z, Eils R, and Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847–2849. [PubMed: 27207943]
- Handel MA 2004. The XY body: a specialized meiotic chromatin domain. *Exp. Cell Res.* 296:57–63. [PubMed: 15120994]
- Hermann BP, Cheng K, Singh A, Roa-De La Cruz L, Mutoji KN, Chen I-C, Gildersleeve H, Lehle JD, Mayo M, Westernströer B, Law NC, Oatley MJ, Velte EK, Niedenberger BA, Fritze D, Silber S, Geyer CB, Oatley JM, and McCarrey JR. 2018. The mammalian spermatogenesis single-cell transcriptome, from spermatogonial stem cells to spermatids. *Cell Rep.* 25:1650–1667.e8.
- Holt J, Huang S, McMillan L, and Wang W. 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, Holt J, Kao C-Y, McMillan L, and Wang W. 2014. A novel multi-alignment pipeline for high-throughput sequencing data. *Database* 2014:bau057.
- Huang S, Kao C-Y, McMillan L, and Wang W. 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, Tsuboi K, Ohishi N, Tsuchiya K, and Matsuda Y. 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. [PubMed: 25801302]
- Kampstra P. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *J. Stat. Softw.* 28.
- Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, and Pääbo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309:1850–1854. [PubMed: 16141373]
- Khaitovich P, Kelso J, Franz H, Visagie J, Giger T, Joerchel S, Petzold E, Green RE, Lachmann M, and Pääbo S. 2006. Functionality of intergenic transcription: an evolutionary comparison. *PLoS Genet.* 2:e171. [PubMed: 17040132]
- Khrameeva E, Kurochkin I, Han D, Guijarro P, Kanton S, Santel M, Qian Z, Rong S, Mazin P, Sabirov M, Bulat M, Efimova O, Tkachev A, Guo S, Sherwood CC, Camp JG, Pääbo S, Treutlein B, and Khaitovich P. 2020. Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res.* 30:776–789. [PubMed: 32424074]

- King MC, and Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116. [PubMed: 1090005]
- Kiselev VY, Andrews TS, and Hemberg M. 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20:273–282. [PubMed: 30617341]
- Kopania EEK, Larson EL, Callahan C, Keeble S, and Good JM. 2021. Molecular evolution across mouse spermatogenesis. *bioRxiv*, doi: 10.1101/2021.08.04.455131.
- Kousathanas A, Halligan DL, and Keightley PD. 2014. Faster-X adaptive protein evolution in house mice. *Genetics* 196:1131–1143. [PubMed: 24361937]
- Laiho A, Kotaja N, Gyenesei A, and Sironen A. 2013. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One* 8:e61558.
- Landeon EL, Muirhead CA, Wright L, Meiklejohn CD, and Presgraves DC. 2016. Sex chromosome-wide transcriptional suppression and compensatory cis-regulatory evolution mediate gene expression in the *Drosophila* male germline. *PLoS Biol.* 14:e1002499.
- Lara NLM, Costa GMJ, Avelar GF, Lacerda SMSN, Hess RA, and de França LR. 2018. Testis physiology—overview and histology. Pp. 105–116 in Skinner MK, ed. *Encyclopedia of Reproduction*. Academic Press, London.
- Larson EL, Keeble S, Vanderpool D, Dean MD, and Good JM. 2017. The composite regulatory basis of the large X-effect in mouse speciation. *Mol. Biol. Evol.* 34:282–295. [PubMed: 27999113]
- Larson EL, Kopania EEK, and Good JM. 2018. Spermatogenesis and the evolution of mammalian sex chromosomes. *Trends Genet.* 34:722–732. [PubMed: 30077434]
- Larson EL, Vanderpool D, Keeble S, Zhou M, Sarver BAJ, Smith AD, Dean MD, and Good JM. 2016. Contrasting levels of molecular evolution on the mouse X chromosome. *Genetics* 203:1841–1857. [PubMed: 27317678]
- Li H, Huang Q, Liu Y, and Garmire LX. 2020. Single cell transcriptome research in human placenta. *Reproduction* 160:R155–R167. [PubMed: 33112783]
- Li Y-F, He W, Jha KN, Klotz K, Kim Y-H, Mandal A, Pulido S, Digilio L, Flickinger CJ, and Herr JC. 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 Sharakhov IV. 2019. Premeiotic and meiotic failures lead to hybrid male sterility in the *Anopheles gambiae* complex. *Proc. Biol. Sci.* 286:20191080.
- Liao Y, Smyth GK, and Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. [PubMed: 24227677]
- Lifschytz E, and Lindsley DL. 1972. The role of X-chromosome inactivation during spermatogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 69:182–186. [PubMed: 4621547]
- Liu Y, Zhou J, and White KP. 2014. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics* 30:301–304. [PubMed: 24319002]
- Lüpold S, Linz GM, Rivers JW, Westneat DF, and Birkhead TR. 2009. Sperm competition selects beyond relative testes size in birds. *Evolution* 63:391–402. [PubMed: 19215291]
- Ma S, Avanesov AS, Porter E, Lee BC, Mariotti M, Zemskaya N, Guigo R, Moskalev AA, and Gladyshev VN. 2018. Comparative transcriptomics across 14 *Drosophila* species reveals signatures of longevity. *Aging Cell* 17:e12740.
- Mack KL, Campbell P, and Nachman MW. 2016. Gene regulation and speciation in house mice. *Genome Res.* 26:451–461. [PubMed: 26833790]
- Mack KL, and Nachman MW. 2017. Gene regulation and speciation. *Trends Genet.* 33:68–80. [PubMed: 27914620]
- Maekawa M, Ito C, Toyama Y, Suzuki-Toyota F, Kimura T, Nakano T, and Toshimori K. 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. [PubMed: 15700541]
- Manceau M, Domingues VS, Mallarino R, and Hoekstra HE. 2011. The developmental role of Agouti in color pattern evolution. *Science* 331:1062–1065. [PubMed: 21350176]

- McCarthy DJ, Chen Y, and Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40:4288–4297. [PubMed: 22287627]
- Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, and Presgraves DC. 2011. Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biol.* 9:e1001126.
- Montgomery SH, and Mank JE. 2016. Inferring regulatory change from gene expression: the confounding effects of tissue scaling. *Mol. Ecol.* 25:5114–5128. [PubMed: 27564408]
- Morgan K, Harr B, White MA, Payseur BA, and Turner LM. 2020. Disrupted gene networks in subfertile hybrid house mice. *Mol. Biol. Evol.* 37:1547–1562. [PubMed: 32076722]
- Mugal CF, Wang M, Backström N, Wheatcroft D, Ålund M, Sémon M, McFarlane SE, Dutoit L, Qvarnström A, and Ellegren H. 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. [PubMed: 33144405]
- Mukaj A, Piálek J, Fotopulosova V, Morgan AP, Odenthal-Hesse L, Parvanov ED, and Forejt J. 2020. Prdm9 intersubspecific interactions in hybrid male sterility of house mouse. *Mol. Biol. Evol.* 37:3423–3438. [PubMed: 32642764]
- Murat F, Mbengue N, Winge SB, Trefzer T, Leushkin E, Sepp M, Cardoso-Moreira M, Schmidt J, Schneider C, Mößinger K, Brüning T, Lamanna F, Belles MR, Conrad C, Kondova I, Bontrop R, Behr R, Khaitovich P, Pääbo S, Marques-Bonet T, Grützner F, Almstrup K, Schierup MH, and Kaessmann H. 2021. The molecular evolution of spermatogenesis across mammals. *bioRxiv*, doi: 10.1101/2021.11.08.467712.
- Namekawa SH, Park PJ, Zhang L-F, Shima JE, McCarrey JR, Griswold MD, and Lee JT. 2006. Postmeiotic sex chromatin in the male germline of mice. *Curr. Biol.* 16:660–667. [PubMed: 16581510]
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, Diehn M, and Alizadeh AA. 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37:773–782. [PubMed: 31061481]
- Nguyen TB, Manova K, Capodici P, Lindon C, Bottega S, Wang X-Y, Refik-Rogers J, Pines J, Wolgemuth DJ, and Koff A. 2002. Characterization and expression of mammalian cyclin b3, a prepachytene meiotic cyclin. *J. Biol. Chem.* 277:41960–41969.
- Oka A, Mita A, Takada Y, Koseki H, and Shiroishi T. 2010. Reproductive isolation in hybrid mice due to spermatogenesis defects at three meiotic stages. *Genetics* 186:339–351. [PubMed: 20610405]
- Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, and Suggests M. 2007. The vegan package. *Community ecology package* 10:719.
- Pitnick S, Hosken DJ, and Birkhead TR. 2009. Sperm morphological diversity. Pp. 69–149 in Birkhead TR, Hosken DJ, and Pitnick S, eds. *Sperm Biology*. Academic Press, London.
- Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, and Wolf JBW. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414. [PubMed: 24948738]
- Poelstra JW, Vijay N, Höppner MP, and Wolf JB. 2015. Transcriptomics of colour patterning and colouration shifts in crows. *Mol. Ecol.* 24:4617–4628. [PubMed: 26302355]
- Price-Waldman RM, Shultz AJ, and Burns KJ. 2020. Speciation rates are correlated with changes in plumage color complexity in the largest family of songbirds. *Evolution* 74:1155–1169. [PubMed: 32333393]
- Rafati N, Blanco-Aguiar JA, Rubin CJ, Sayyab S, Sabatino SJ, Afonso S, Feng C, Alves PC, Villafuerte R, Ferrand N, Andersson L, and Carneiro M. 2018. A genomic map of clinal variation across the European rabbit hybrid zone. *Mol. Ecol.* 27:1457–1478. [PubMed: 29359877]
- Ramm SA, and Schärer L. 2014. The evolutionary ecology of testicular function: size isn't everything. *Biol. Rev. Camb. Philos. Soc.* 89:874–888. [PubMed: 24495304]
- Ramm SA, Schärer L, Ehmcke J, and Wistuba J. 2014. Sperm competition and the evolution of spermatogenesis. *Mol. Hum. Reprod.* 20:1169–1179. [PubMed: 25323971]

- Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, and Zarkower D. 2000. Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev.* 14:2587–2595. [PubMed: 11040213]
- Robinson MD, McCarthy DJ, and Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. [PubMed: 19910308]
- Rohlf RV, and Nielsen R. 2015. Phylogenetic ANOVA: the expression variance and evolution model for quantitative trait evolution. *Syst. Biol.* 64:695–708. [PubMed: 26169525]
- Rolland AD, Lareyre J-J, Goupil A-S, Montfort J, Ricordel M-J, Esquerré D, Hugot K, Houlgatte R, Chalmel F, and Le Gac F. 2009. Expression profiling of rainbow trout testis development identifies evolutionary conserved genes involved in spermatogenesis. *BMC Genomics* 10:546. [PubMed: 19925684]
- Sato DX, Rafati N, Ring H, Younis S, Feng C, Blanco-Aguilar JA, Rubin C-J, Villafuerte R, Hallböök F, Carneiro M, and Andersson L. 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. [PubMed: 32835359]
- Schärer L, Littlewood DTJ, Waeschenbach A, Yoshida W, and Vizoso DB. 2011. Mating behavior and the evolution of sperm design. *Proc. Natl. Acad. Sci. U. S. A.* 108:1490–1495. [PubMed: 21220334]
- Schultz N, Hamra FK, and Garbers DL. 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 DJ, Wang RJ, White MA, and Payseur BA. 2018. Genetic dissection of hybrid male sterility across stages of spermatogenesis. *Genetics* 210:1453–1465. [PubMed: 30333190]
- Shen-Orr SS, and Gaujoux R. 2013. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr. Opin. Immunol.* 25:571–578. [PubMed: 24148234]
- Shima JE, McLean DJ, McCarrey JR, and Griswold MD. 2004. The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol. Reprod.* 71:319–330. [PubMed: 15028632]
- Snyder RL 1967. Fertility and reproductive performance of grouped male mice. Pp. 458–472 in K. Benirschke, ed. *Comparative aspects of reproductive failure*. Springer, Berlin.
- Srivatsan SR, Regier MC, Barkan E, Franks JM, Packer JS, Grosjean P, Duran M, Saxton S, Ladd JJ, Spielmann M, Lois C, Lampe PD, Shendure J, Stevens KR, and Trapnell C. 2021. Embryo-scale, single-cell spatial transcriptomics. *Science* 373:111–117. [PubMed: 34210887]
- Stern DL, and Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–2177. [PubMed: 18616572]
- Suzuki TA, and Nachman MW. 2015. Speciation and reduced hybrid female fertility in house mice. *Evolution* 69:2468–2481. [PubMed: 26299202]
- Todd EV, Black MA, and Gemmell NJ. 2016. The power and promise of RNA-seq in ecology and evolution. *Mol. Ecol.* 25:1224–1241. [PubMed: 26756714]
- Torgasheva AA, and Borodin PM. 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 DG, Kulathinal RJ, and Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol. Biol. Evol.* 19:1973–1980. [PubMed: 12411606]
- Trapnell C, Pachter L, and Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111. [PubMed: 19289445]
- Turner LM, Chuong EB, and Hoekstra HE. 2008. Comparative analysis of testis protein evolution in rodents. *Genetics* 179:2075–2089. [PubMed: 18689890]
- Turner LM, and Harr B. 2014. Genome-wide mapping in a house mouse hybrid zone reveals hybrid sterility loci and Dobzhansky-Muller interactions. *Elife* 3:e02504.

- Turner LM, Schwahn DJ, and Harr B. 2012. Reduced male fertility is common but highly variable in form and severity in a natural house mouse hybrid zone. *Evolution* 66:443–458. [PubMed: 22276540]
- Turner LM, White MA, Tautz D, and Payseur BA. 2014. Genomic networks of hybrid sterility. *PLoS Genet.* 10:e1004162.
- Wang Z, Gerstein M, and Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10:57–63. [PubMed: 19015660]
- White-Cooper H, Doggett K, and Ellis RE. 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, Karr TL, Mongue AJ, Dorus S, and Walters JR. 2019. Evolutionary proteomics reveals distinct patterns of complexity and divergence between Lepidopteran sperm morphs. *Genome Biol. Evol.* 11:1838–1846. [PubMed: 31268533]
- Widmayer SJ, Handel MA, and Aylor DL. 2020. Age and genetic background modify hybrid male sterility in house mice. *Genetics* 216:585–597. [PubMed: 32817010]
- Wittkopp PJ 2007. Variable gene expression in eukaryotes: a network perspective. *J. Exp. Biol.* 210:1567–1575. [PubMed: 17449821]
- Wohnhaas CT, Leparac GG, Fernandez-Albert F, Kind D, Gantner F, Viollet C, Hildebrandt T, and Baum P. 2019. DMSO cryopreservation is the method of choice to preserve cells for droplet-based single-cell RNA sequencing. *Sci. Rep.* 9:10699.
- Yapar E, Saglican E, Dönerta HM, Özkurt E, Yan Z, Hu H, Guo S, Erdem B, Rohlfsv RV, Khaitovich P, and Somel M. 2021. Convergent evolution of primate testis transcriptomes reflects mating strategy. *bioRxiv*, doi: 10.1101/010553.
- Yoshikawa H, Xu D, Ino Y, Yoshino T, Hayashida T, Wang J, Yazawa R, Yoshizaki G, and Takeuchi Y. 2018. Hybrid sterility in fish caused by mitotic arrest of primordial germ cells. *Genetics* 209:507–521. [PubMed: 29610216]

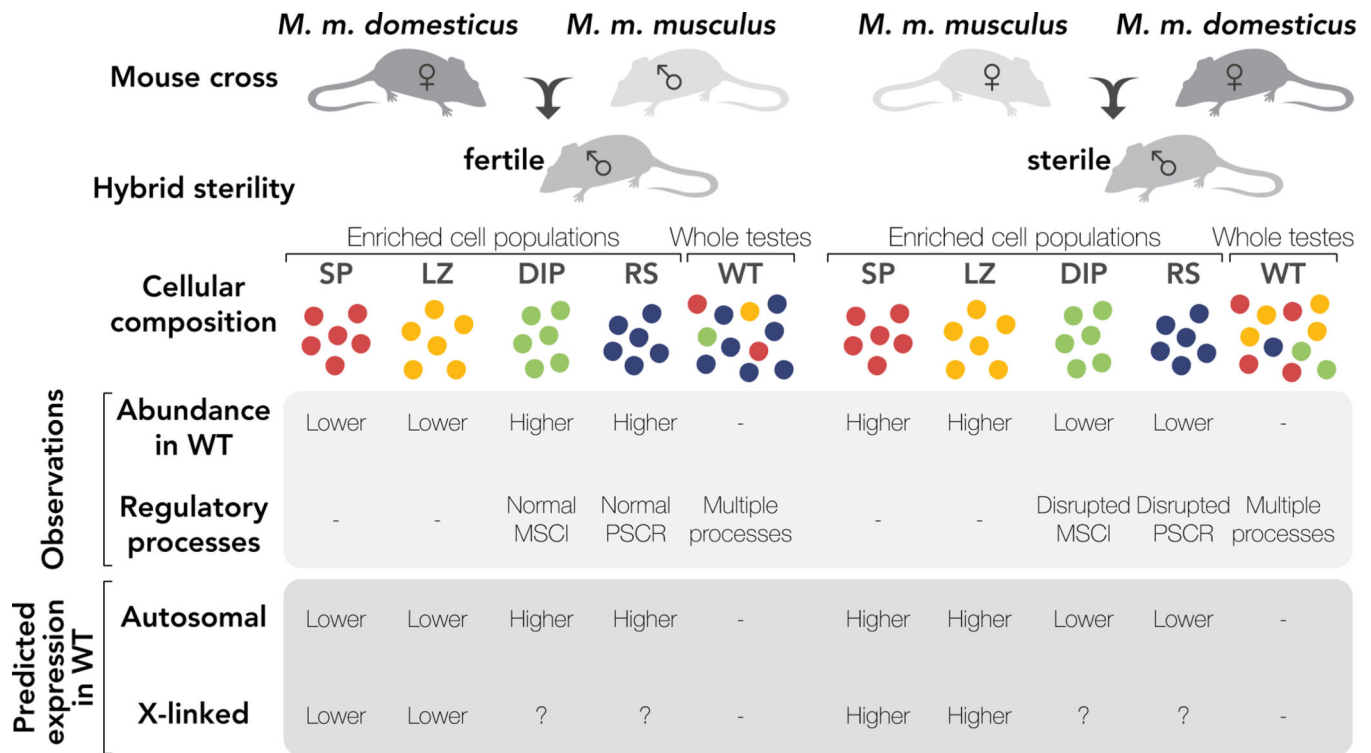


Fig. 1. Crossing design, sampling approach, and predicted cellular composition and expression differences between reciprocal hybrids.

We compared expression patterns using two sampling approaches, enriched cell populations (red = Mitosis (SP), yellow = Meiosis^{Before X-Inact.} (LZ), green = Meiosis^{After X-Inact.} (DIP), blue = Postmeiosis (RS)) and Whole Testes (WT). Whole testes are susceptible to changes in cellular composition between hybrids while enriched cell populations should be buffered from these effects. Relative expression of autosomal stage-specific genes from each enriched cell population in whole testes samples is predicted to track changes in cellular composition between *sterile* and *fertile* mice. Relative expression of X-linked stage-specific genes from enriched cell populations is predicted to be influenced by both changes in cellular composition and expected regulatory processes operating in those cell types.

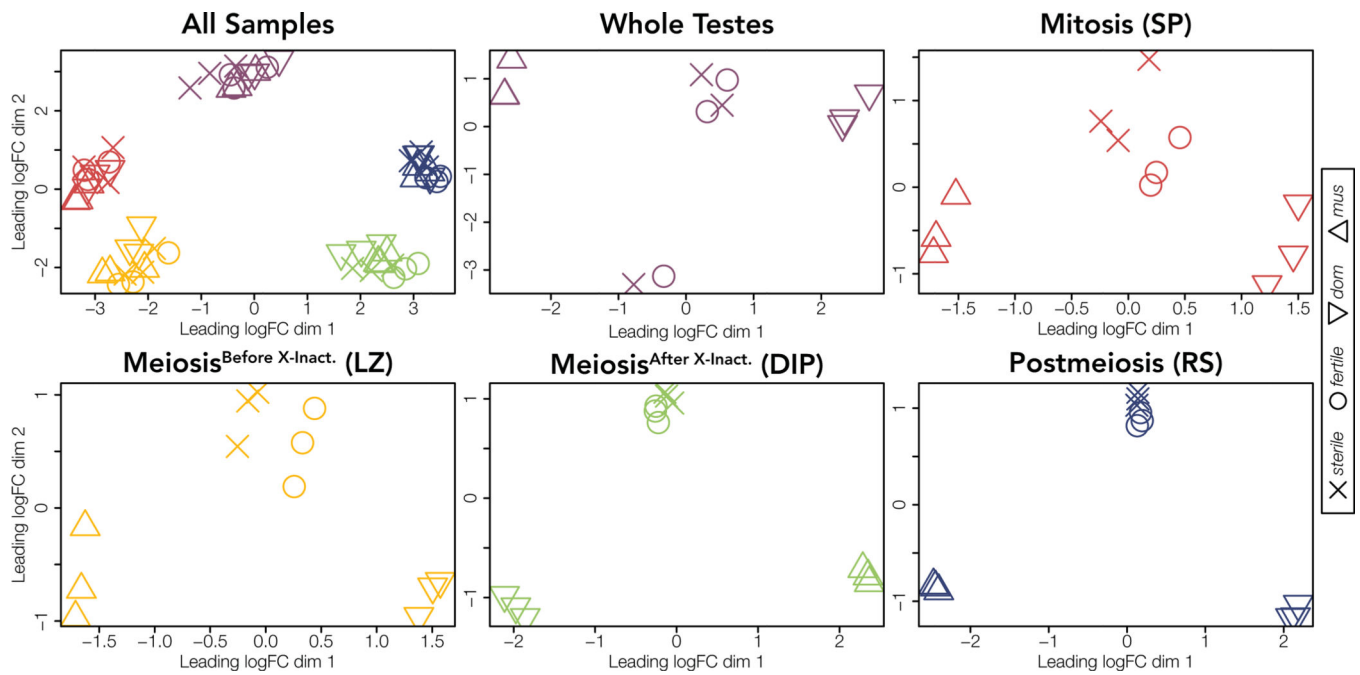


Figure 2. Sample type then cross type drives differences in expression profiles.

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 \log_2 fold changes among genes that distinguish each sample. Each cross is indicated by a symbol (*mus* = Δ , *dom* = ∇ , *fertile* = O, and *sterile* = X). Samples are colored by sample type (red = Mitosis, yellow = Meiosis^{Before X-Inact.}, green = Meiosis^{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. Each sample type for each cross is represented by three replicates.

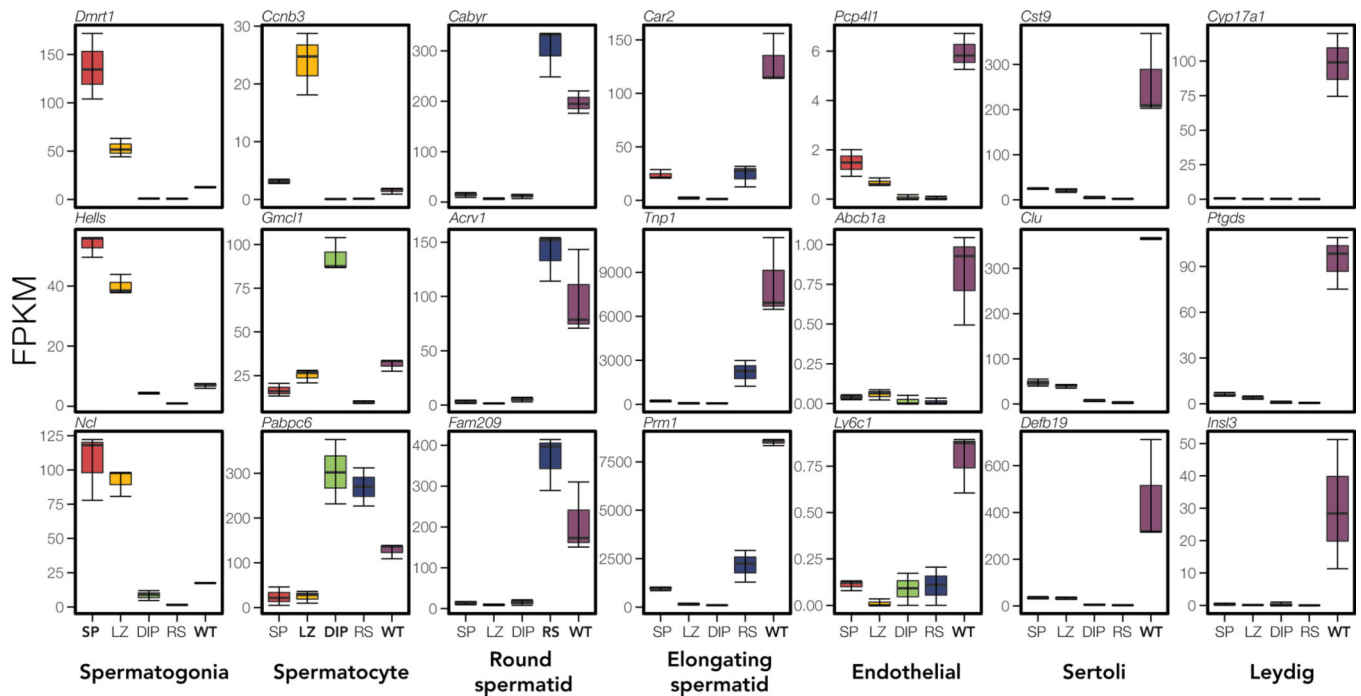


Figure 3. Whole testes show signatures of more diverse cell types than enriched cell populations. Expression of cell-specific marker genes (Green et al. 2018) across each sample type for *dom* reference samples. We quantified expression (FPKM) of three marker genes (rows) associated with testes-specific cell types (columns). Each panel displays marker expression in each sample type (red = Mitosis (SP), yellow = Meiosis^{Before X-Inact.} (LZ), green = Meiosis^{After X-Inact.} (DIP), blue = Postmeiosis (RS), and purple = Whole Testes (WT)). Sample types are bolded in each panel where marker gene expression is expected. Note, *Ccnb3* expression is specific to Meiotic^{Before X-Inact.} cells (Maekawa et al. 2004), and *Gmcl1* is specific to Meiotic^{After X-Inact.} cells (Nguyen et al. 2002).

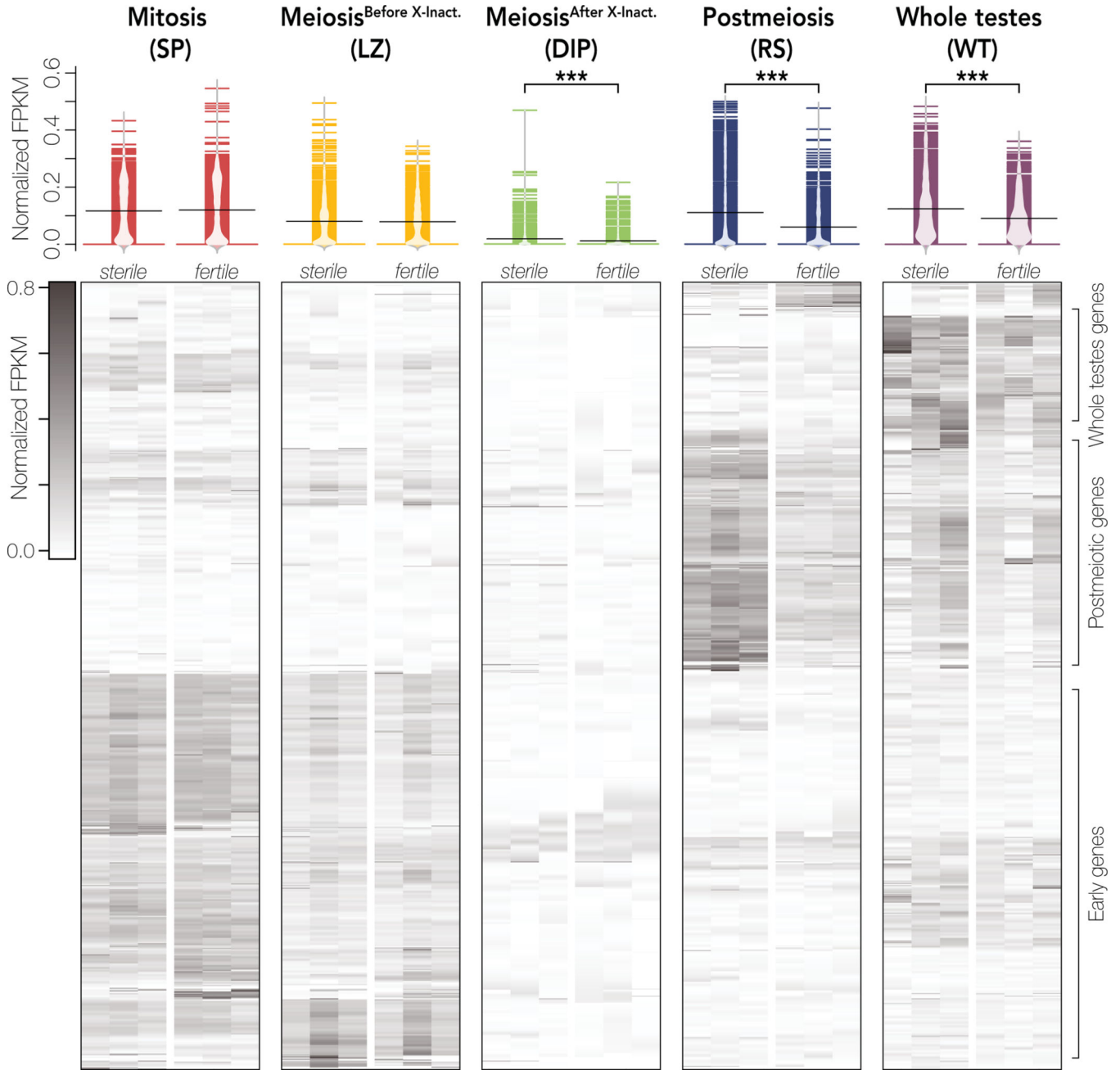


Figure 4. Patterns of X-linked gene expression in *sterile* and *fertile* hybrids differ between sorted cells and whole testes.

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). Expression distributions are colored by sample type (red = Mitosis, yellow = Meiosis^{Before X-Inact.}, green = Meiosis^{After X-Inact.}, blue = Postmeiosis, and purple = Whole Testes) and labelled by cross (*sterile* or *fertile* hybrid). These plots were generated with the R package *beanplot* (Kampstra 2008) and differences in expression were calculated with Wilcoxon signed-rank tests where *** indicates $p < 0.001$ and * indicates $p < 0.05$ after FDR correction

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(Benjamini and Hochberg 1995). 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 represents a gene with darker colors indicating higher expression. The heatmap was generated with the R package ComplexHeatmap v.2.3.2 (Gu *et al.* 2016).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

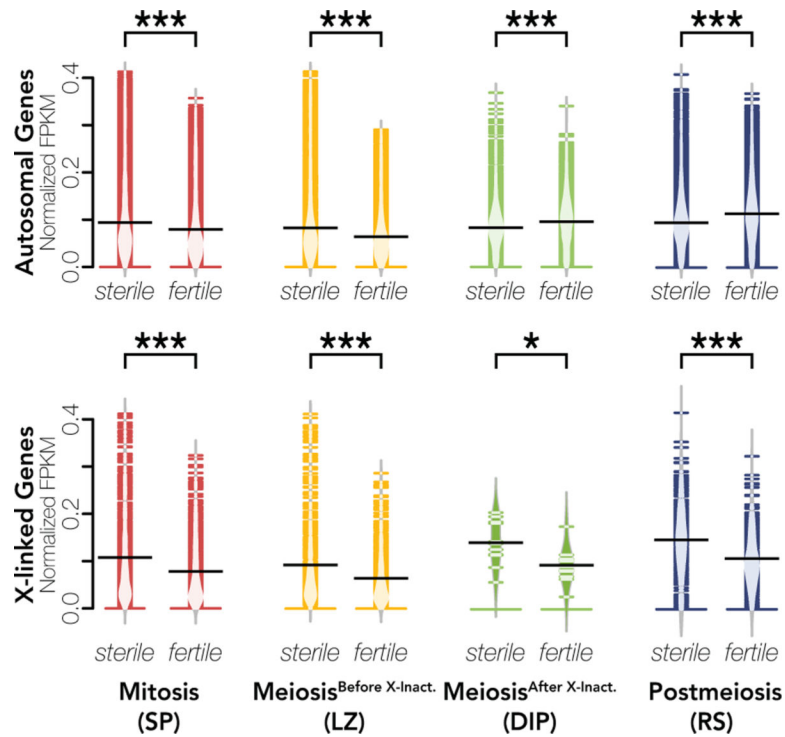


Figure 5. Changes in cellular composition alter expression of stage-specific genes in whole testes samples.

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^{Before X-Inact.} cells are present at lower abundances in *sterile* hybrids while Meiotic^{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).

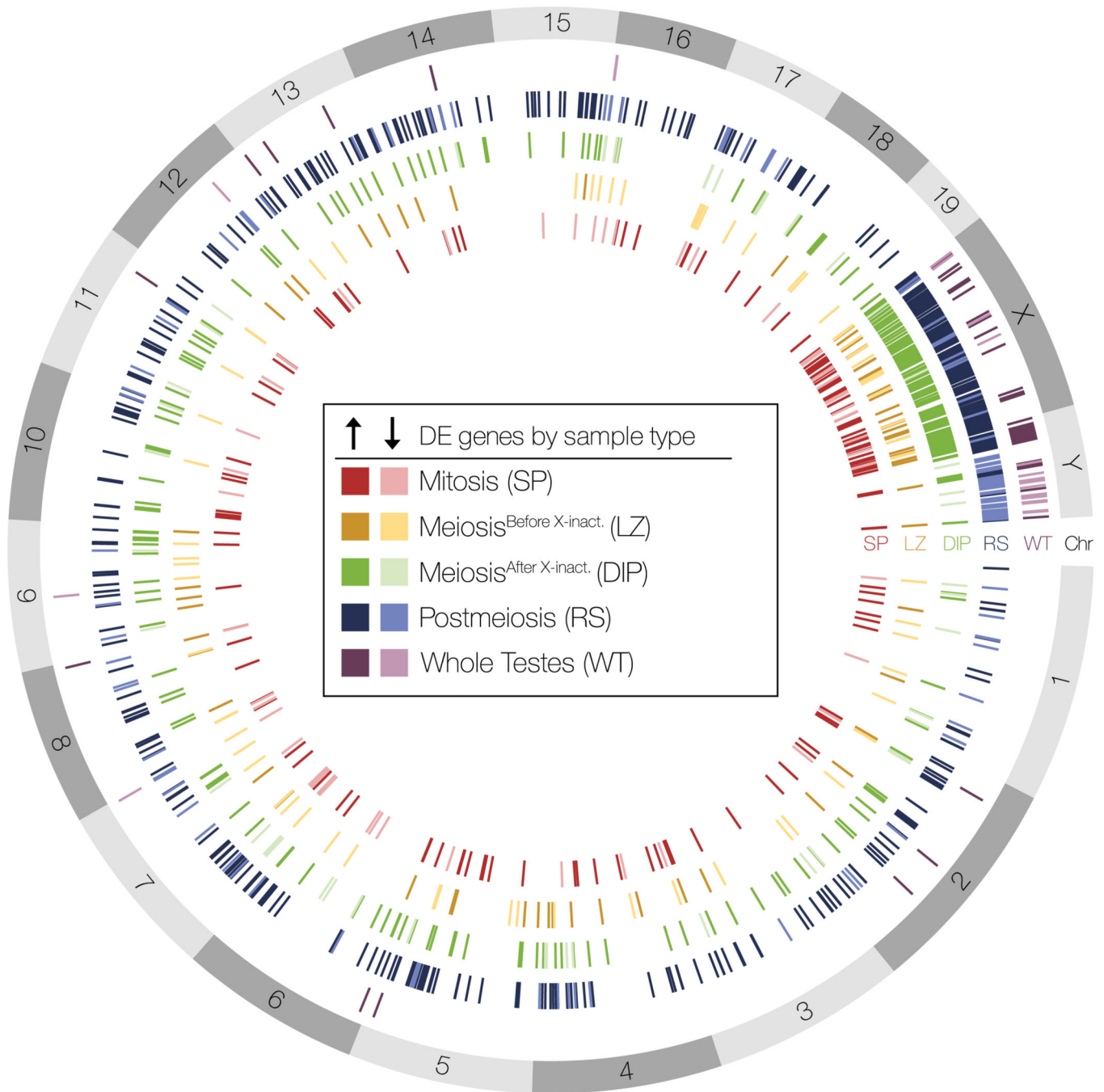


Figure 6. Whole testes and enriched cell populations differed in the number and identity of differentially expressed genes.

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.

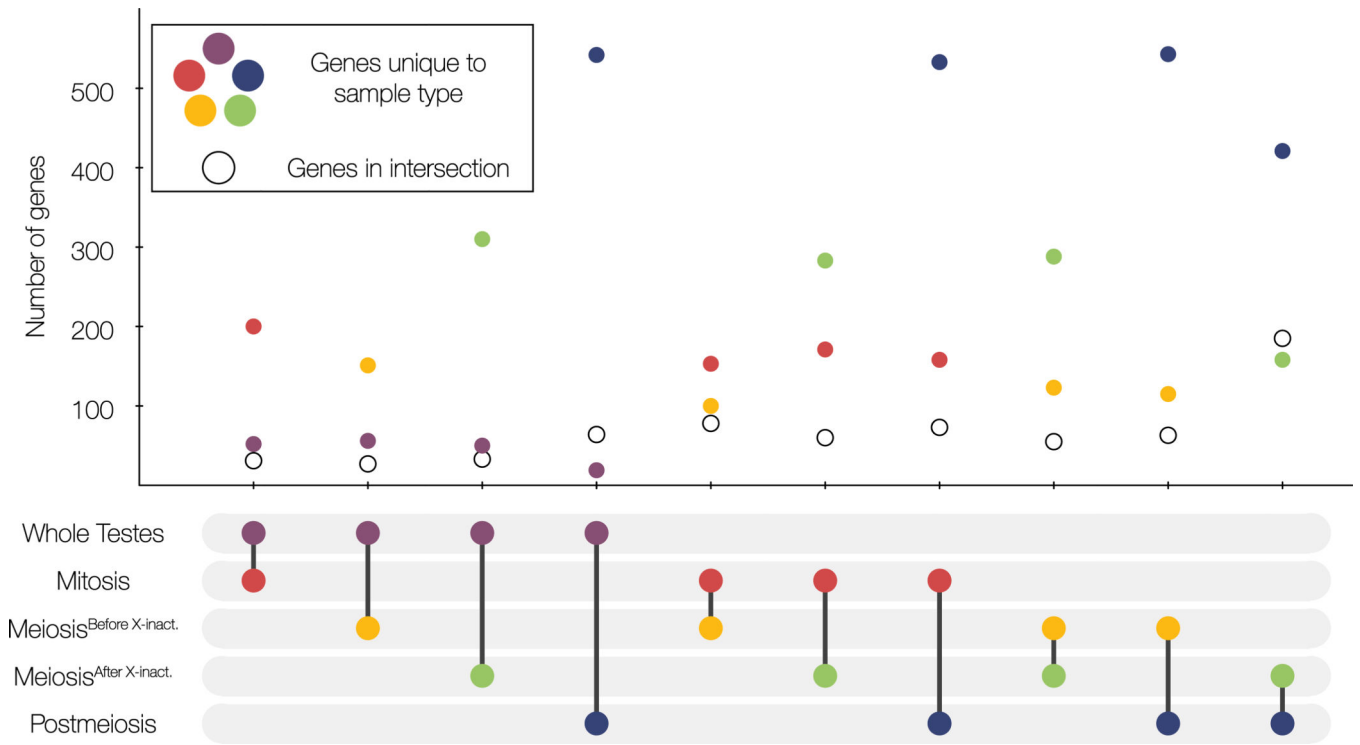


Figure 7. Whole testes and enriched cell populations differed in pairwise comparisons of DE genes between *sterile* and *fertile* hybrids.

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^{Before X-Inact.}, green = Meiosis^{After X-Inact.}, blue = Postmeiosis, and purple = Whole Testes).