Title: Unraveling patterns of disrupted gene expression across a complex tissue

Authors: 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

47 INTRODUCTION

48 A single genome acts as the blueprint for all of the diverse cell types that comprise a eukaryotic
49 organism. This diversity of cellular function is achieved through the expression of individual
50 genes orchestrated by large, layered regulatory networks (Davidson and Erwin 2006; Wittkopp
51 2007). Often it is through gene expression that changes to the genome are connected to higher
52 level organismal phenotypes of primary interest, and the evolution of gene expression itself can
53 profoundly influence a species' evolutionary trajectory (King and Wilson 1975; Carroll 2008;
54 Stern and Orgogozo 2008). Gene expression is not a static biochemical phenotype – it is an
55 amalgamation of expression profiles of individual cell types as genes are turned on and off
56 across organismal space and developmental time. Bulk RNASeq of whole tissues allows us to
57 investigate these dynamics in non-model systems with minimal genomic resources and is
58 affordable and tractable for field-based studies (Alvarez et al. 2015). However, evolutionarily
59 important phenotypes often manifest in complex heterogenous tissues, such as sterility in
60 reproductive organs (Turner *et al.* 2012; Suzuki and Nachman 2015), behavioral changes in
61 neurological tissue (Sato *et al.* 2020), or color patterning across the body (Manceau *et al.* 2011;
62 Poelstra *et al.* 2014). Standard bulk sequencing approaches necessarily collapse the complexity
63 inherent to gene expression in these tissues and implicitly assume equivalent proportions of cell
64 types across different comparisons. But if the relative abundance of cell types differs between
65 contrasts, then we may be unable to distinguish regulatory divergence from differences in
66 cellular composition (Good et al. 2010; Montgomery and Mank 2016). What are the
67 consequences of using a whole tissue approach on expression profiles and how does this impact
68 inferences on evolutionary divergence?

69 Testes are emblematic of a complex tissue and are central to reproductive divergence and
70 speciation. Testes genes are among the most rapidly evolving at the level of protein sequence
71 (Torgerson et al. 2002; Good and Nachman 2005; Turner et al. 2008; Larson et al. 2016) and
72 gene expression (Brawand *et al.* 2011). Sperm, which are produced by the testes, are among the
73 most morphologically diverse animal cells (Pitnick et al. 2009) and are critical in both prezygotic
74 (*e.g.,* sperm competition) and postzygotic (*e.g.,* hybrid sterility) reproductive barriers between
75 species. Studies of whole testes expression have yielded great insights into the evolution of male
76 reproductive traits (*e.g.,* Catron and Noor 2008; Davis et al. 2015; Mack et al. 2016; Ma et al.
77 2018; Rafati et al. 2018), but relatively few studies have accounted for the cellular complexity of
78 testes, a factor which we expect to complicate evolutionary inference from whole tissues (Good
79 et al. 2010). Testes are dominated by various stages of developing sperm, primarily postmeiotic
80 cells (~ 70% in house mice; Bellvé et al. 1977), but also present are mitotic precursors,
81 endothelial cells, support cells (White-Cooper et al. 2009), and even multiple types of sperm in
82 some organisms (Whittington *et al.* 2019). The relative proportion of testes cell types is
83 evolvable and plastic (Ramm and Schärer 2014; Ramm et al. 2014) and can vary across species
84 (Lara *et al.* 2018), mating strategies (Firman *et al.* 2015), age (Ernst *et al.* 2019; Widmayer *et al.*
85 2020), and social conditions (Snyder 1967). For all these reasons, we might expect the cellular
86 composition of testes to differ – sometimes dramatically – between different species,
87 populations, or experimental contrasts.

88 The cellular complexity of tissues is often due to the developmental complexity of the
89 phenotypes those tissues produce. In testes, undifferentiated germ cells (spermatogonia) undergo
90 multiple rounds of mitosis then enter meiosis (spermatocytes) where they undergo two rounds of
91 cell division to produce four haploid cells (round spermatids). These cells then undergo dramatic

92 postmeiotic differentiation to produce mature spermatozoa. Each of these stages has a unique
93 gene expression profile (Shima et al. 2004; Green et al. 2018; Hermann et al. 2018) and is
94 subject to different selective pressures (Larson et al. 2018). Spermatogenesis in many animals
95 has an additional layer of developmental complexity in the form of the intricate regulation of the
96 sex chromosomes. During early meiosis in mice, the X chromosome is completely
97 transcriptionally inactivated (meiotic sex chromosome inactivation or MSCI; Handel 2004) and
98 remains repressed for the remainder of spermatogenesis (postmeiotic sex chromosome repression
99 or PSCR; Namekawa et al. 2006). Bulk whole testes sequencing aggregates these diverse
100 developmental stages, limiting our resolution into how the molecular mechanisms underlying
101 phenotypic change act in a developmental context (Larson et al. 2018).

102 The combination of the cellular heterogeneity and developmental complexity of testes is
103 particularly relevant in understanding the evolution of hybrid male sterility. We expect sterile
104 hybrids to have disrupted testes expression (Mack and Nachman 2017; Morgan *et al.* 2020), but
105 sterile hybrids are also likely to have different testes cell composition compared to fertile mice.
106 For example, some sterile house mouse hybrids have only a fourth as many postmeiotic cells
107 (Schwahn *et al.* 2018). These differences in cell composition alone might cause what looks like
108 differential gene regulation associated with hybridization. This is especially problematic when
109 differences in cell composition correspond to developmental timepoints where hybrid expression
110 is disrupted, such as with the disruption of X chromosome inactivation at MSCI (Good et al.
111 2010; Bhattacharyya et al. 2013; Campbell et al. 2013; Larson et al. 2017), and it is not clear
112 how patterns of stage-specific disruption in hybrids appear in whole testes where stages
113 exhibiting normal and disrupted X regulation are combined. Evidence for disrupted X
114 chromosome regulation in sterile hybrids varies across taxa (Davis et al. 2015; Rafati et al. 2018;

115 Bredemeyer et al. 2021), but outside of mice, most studies have been restricted to whole testes
116 RNASeq. Although these potentially confounding factors are often acknowledged in whole
117 tissue studies (Good et al. 2010; Turner et al. 2014; Davis et al. 2015; Mugal et al. 2020), no
118 systematic effort has been made to distinguish how differences in cellular composition can be
119 distinguished from underlying regulatory dynamics in hybrids using whole testes samples.

120 Here, we use two analogous RNASeq datasets of fertile and sterile F1 hybrids from *Mus*
121 *musculus musculus* and *M. m. domesticus* house mice (Mack *et al.* 2016; Larson *et al.* 2017) as a
122 model to investigate the effects of bulk whole tissue sequencing on divergent gene expression.
123 These subspecies form a hybrid zone in Europe where they produce subfertile hybrid males
124 (Turner et al. 2012). F1 hybrid males from wild-derived strains differ in severity of sterility
125 dependent on the strains and the direction of the cross (Britton-Davidian et al. 2005; Good et al.
126 2008; Mukaj et al. 2020), with more sterile crosses having greatly disrupted cellular composition
127 and gene expression (Good et al. 2010; Bhattacharyya et al. 2013; Campbell et al. 2013; Turner
128 and Harr 2014; Larson et al. 2017; Schwahn et al. 2018). We use comparisons of fertile and
129 sterile reciprocal F1 hybrids to disentangle the effects of cellular composition and disrupted
130 regulatory processes on divergent gene expression. We first examine which cell types contribute
131 to whole testes expression profiles then test predictions about the effects of cell type abundance
132 on whole testes comparisons. Finally, we assess whether signatures of disrupted gene regulation
133 during specific stages of spermatogenesis are detectable in a whole tissue approach and the
134 consequences of whole tissue sampling on differential gene expression. Collectively, we show
135 that inferences from comparative bulk RNASeq approaches are sensitive to changes in cellular
136 composition in complex tissues and advocate for an increased awareness of histology and tissue
137 morphology during study design of RNASeq in non-model systems to account for such effects.

138 MATERIALS AND METHODS

139 *Mouse strains and datasets*

140 We used gene expression data from two recently published datasets analyzing disrupted hybrid
141 gene expression in whole testes (SRA PRJNA286765; Mack et al. 2016) and enriched cell
142 populations across four stages of spermatogenesis (SRA PRJNA296926; Larson et al. 2017).
143 Both studies sequenced transcriptomes from the same wild-derived inbred strains of the mouse
144 subspecies *M. m. domesticus* and *M. m. musculus*, and their F1 hybrids. For each subspecies, two
145 strains were crossed to generate intraspecific F1s to serve as parental controls, without the effects
146 of inbreeding depression on fertility (Good *et al.* 2008). The *M. m. domesticus* mice were
147 generated by crossing the strains WSB/EiJ and LEWES/EiJ, with LEWES dams for the whole
148 testes dataset and WSB dams for the enriched cell dataset (hereafter *dom*). *M. m. musculus* mice
149 were generated by crossing the strains PWK/PhJ and CZECHII/EiJ with PWK dams for the
150 whole testes dataset and CZECHII dams for the sorted cell dataset (hereafter *mus*). F1 hybrid
151 mice with differing severity of sterility were generated by reciprocally crossing LEWES and
152 PWK; PWK female × LEWES male hybrids are mostly sterile (hereafter *sterile*), LEWES female
153 × PWK male hybrids are mostly fertile (hereafter *fertile*). Mack et al. (2016) produced RNASeq
154 libraries from whole testes for each of the four crosses ((2 parental crosses + 2 hybrid crosses) x
155 3 replicates per cross, N = 12). Larson et al. (2017) used Fluorescence-Activated Cell Sorting
156 (FACS) to isolate enriched cell populations from four different stages of spermatogenesis:
157 Mitosis: spermatogonia (SP), Meiosis$^{\text{Before X-Inact.}}$: leptotene and zygotene spermatocytes (LZ),
158 Meiosis$^{\text{After X-Inact.}}$: diplotene spermatocytes (DIP), and Postmeiosis: round spermatids (RS) ((2
159 parental crosses + 2 hybrid crosses) x 3 replicates per cross x 4 cell types per replicate, N = 48).
160 In both studies, libraries were sequenced on an Illumina HiSeq 2000 (100 bp, PE).

161 *Read mapping and count estimation*

162 We processed both datasets in parallel. First, we used Trimmomatic v.0.38 (Bolger *et al.*
163 2014) to trim low quality bases from the first and last 5 bp of each read and bases averaging a
164 Phred score of less than 15 across a 4 bp sliding window. We retained reads with a minimum
165 length of 36 bp (Table S1). To avoid mapping bias, we aligned trimmed reads to published
166 pseudo-reference genomes for *M. m. musculus* and *M. m. domesticus* (Huang et al. 2007) using
167 TopHat v.2.1.1 (Trapnell *et al.* 2009) and retained up to 250 alignments per read for multi-
168 mapped reads. We used Lapels v.1.1.1 to convert alignments to the reference mouse genome
169 coordinates (GRCm38.p6) and merged alignments with suspenders v.0.2.6 (Holt et al. 2013;
170 Huang et al. 2014). We summarized read counts for annotated genes (Ensembl Release 96) using
171 FeatureCounts v.1.4.4 (Liao *et al.* 2014) for read pairs that aligned to the same chromosome (-B
172 and -C). We analyzed the count data with and without multi-mapped reads (-M) and across all
173 annotated genes or protein-coding genes only. We found consistent results using all approaches
174 and here present results using only uniquely mapped reads for all annotated genes, unless
175 otherwise specified. Whole testes samples averaged ~24 million mapped read pairs per sample
176 while sorted cell populations averaged ~8 million read pairs.

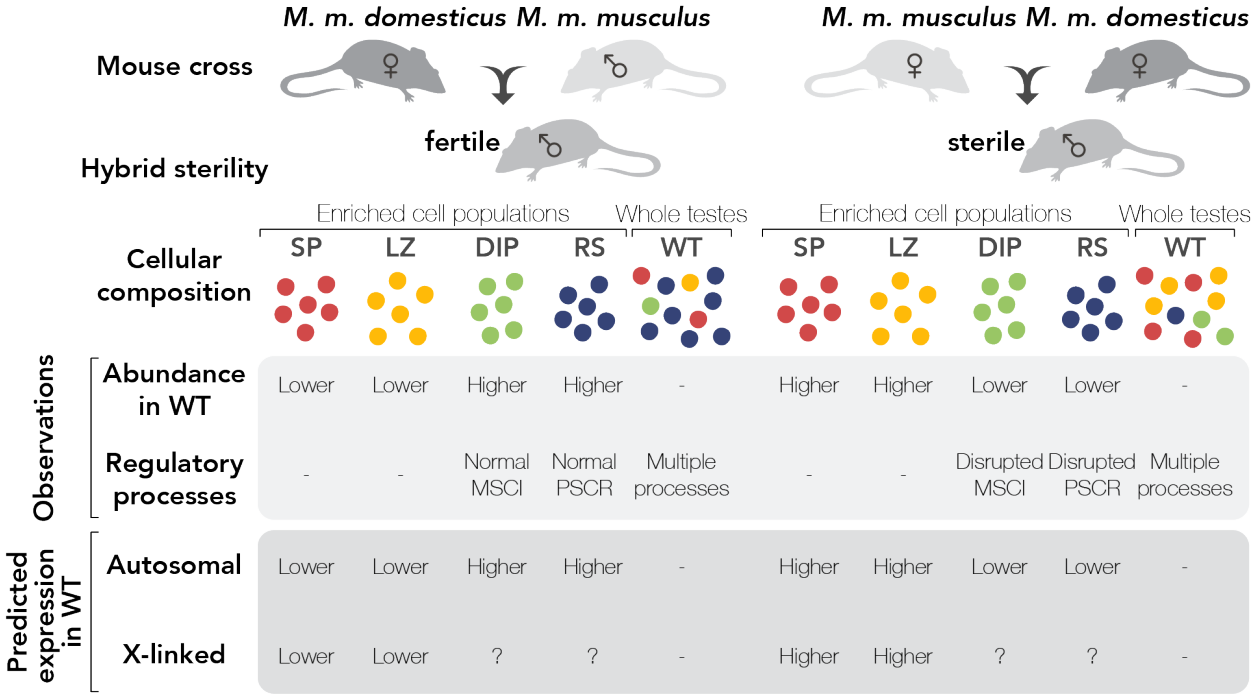
177

178 *Characterizing expression patterns*

179 To investigate how expression differed between both datasets, we defined expressed
180 genes as those with a minimum of one Fragment Per Kilobase of exon per Million mapped reads
181 (FPKM) in at least 3 samples within each dataset. This restricted our analysis to 16,824 genes
182 (12,587 protein-coding) in the whole testes dataset and 21,762 genes (14,284 protein-coding) in
183 the sorted cell dataset. We used R v.4.0.2 for all analyses. We conducted expression analyses

184 using the Bioconductor v.3.11 package edgeR v.3.30.3 (Robinson *et al.* 2010) and normalized
185 the data using the scaling factor method (Anders and Huber 2010).
186
187 *Effects of cellular composition on whole testes expression*
188 To first determine which cell types were present and contributing to the expression
189 profiles of both datasets, we tested all sample types for the expression of marker genes known to
190 be specifically expressed in certain cell types. We selected three marker genes from seven testes
191 cell types: spermatogonia, spermatocytes, round spermatids, elongating spermatids, Sertoli cells,
192 epithelial cells, and Leydig cells (Raymond *et al.* 2000; Nguyen *et al.* 2002; Maekawa *et al.*
193 2004; Li *et al.* 2007; Green *et al.* 2018) as well as marker genes from Hermann et al. (2018).
194 This approach allowed us to assess the purity of sorted cell populations by looking for the
195 expression of non-target cell types in sorted cell populations. We were also able to identify
196 which cell types contributed to the expression profile of whole testes.
197 Next, we tested the hypothesis that differential expression of stage-specific genes in
198 whole tissues can be caused by differences in the relative abundance of cell types between
199 comparisons—in this case *sterile* and *fertile* F1 hybrids (Fig 1; Good et al. 2010). We defined
200 sets of stage-specific genes using our sorted cell populations of each subspecies (Figs S1A, B;
201 Supplemental File 1). We considered a gene to be specific to a given cell population if its median
202 expression (normalized FPKM) was greater than two times its median expression across all other
203 sorted cell populations (*i.e.,* an induced gene approach as in Kousathanas et al. 2014). We then
204 compared the expression of these stage-specific genes in whole testes of *sterile* and *fertile*
205 hybrids. We did this separately for autosomal and X-linked genes because we expected the
206 forces driving patterns of expression to differ between the two. For autosomal genes, we

207 expected expression to be driven largely by differences in cell composition (*e.g.,* fewer later-
208 stage cell types in *sterile* hybrids should lead to lower expression of stage-specific genes from
209 later stages in *sterile* compared to *fertile* whole testes). In contrast, X chromosome inactivation is
210 disrupted in *sterile* hybrids, which should lead to higher expression of stage-specific genes from
211 later stages in *sterile* whole testes. For autosomal genes, we used one-sided paired Wilcoxon
212 signed-rank tests to test if expression of stage-specific genes from more abundant cell types
213 (Mitosis and Meiosis$^{\text{Before X-Inact.}}$) was greater in *sterile* hybrid whole testes and if expression of
214 stage-specific genes from less abundant cell types (Meiosis$^{\text{After X-Inact.}}$ and Postmeiosis) was lower
215 in *sterile* hybrid whole testes. Because we did not know whether the effects of differing cellular
216 compositions or misregulation of the X chromosome would be stronger for driving expression
217 patterns of stage-specific X-linked genes in whole testes, we used two-sided Wilcoxon signed-
218 rank tests for X-linked genes.



219

220 **Fig. 1. Crossing design, sampling approach, and predicted cell composition and expression**
221 **differences between reciprocal hybrids.** We compared expression patterns using two sampling

222 approaches, enriched cell populations (red = Mitosis (SP), yellow = Meiosis$^{\text{Before X-Inact.}}$ (LZ),
223 green = Meiosis$^{\text{After X-Inact.}}$ (DIP), blue = Postmeiosis (RS)) and Whole Testes (WT). Whole testes
224 are susceptible to changes in cellular composition between hybrids while enriched cell
225 populations should be buffered from these effects. Relative expression of autosomal stage-
226 specific genes from each enriched cell population in whole testes samples is predicted to track
227 changes in cellular composition between *sterile* and *fertile* mice. Relative expression of X-linked
228 stage-specific genes from enriched cell populations is predicted to be influenced by both changes
229 in cellular composition and expected regulatory processes operating in those cell types.
230
231 To look for additional signatures of disrupted X-linked gene expression in both sampling
232 approaches, we first used one-sided Wilcoxon signed-rank tests to compare expression of X-
233 linked genes in *sterile* compared to *fertile* hybrids for each sample type where disrupted X-linked
234 expression was expected (*i.e.,* Meiosis$^{\text{After X-Inact.}}$, Postmeiosis, and Whole Testes). Then, we
235 defined sets of "detected" genes for each sample type as those expressed above an FPKM
236 threshold within a replicate (FPKM > 0 – 4) and ran one-way ANOVAs on the number of
237 detected X-linked genes in each cross within a sample type and conducted posthoc Tukey's tests.
238 Note, because there are only three replicates per sample type, these data inevitability violate
239 distribution assumptions in both parametric and non-parametric tests, and differences among
240 treatments should be considered largely qualitative.
241
242 *Differential expression analysis*
243 We conducted differential expression analysis between *sterile* and *fertile* hybrids for all
244 five sample types in edgeR. We fit each dataset (whole testes and sorted cells separately) with

245 negative binomial generalized linear models with Cox-Reid tagwise dispersion estimates
246 (McCarthy *et al.* 2012) and adjusted *P*-values to a false discovery rate (FDR) of 5% (Benjamini
247 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). We used two bootstrapping
251 approaches to determine whether BCVs differed across datasets. First, we subsampled raw count
252 files for 10000 genes across 100 replicates and recalculated the BCV for four groups: hybrid
253 whole testes, parent whole testes, hybrid sorted cells, and parent sorted cells. Second, we
254 dropped one individual per group and recalculated the BCV for every set of n-1 individuals. For
255 both approaches, we estimated 99% confidence intervals for the bootstrap BCV estimates from
256 each group and approach (reported as CI_1 and CI_2, respectively).

257 We contrasted expression between *sterile* and *fertile* hybrids so that a positive log fold-
258 change (logFC) indicated over-expression in *sterile* males. For all pairwise comparisons of
259 sample types, we assessed the number of genes overlapping between both sets of differentially
260 expressed (DE) genes and the number of DE genes unique to each sample type in the
261 comparison. We also calculated whether the direction of fold change for a particular DE gene
262 switched between sample types (*e.g.,* an up-regulated DE gene in *sterile* whole testes that was a
263 down-regulated DE gene in any of the *sterile* sorted cell populations). We extended this analysis
264 comparing the direction of DE genes between sample types to parental samples, contrasting
265 expression between *mus* and *dom* parents so that a positive logFC indicated over-expression in
266 *mus* males. We tested for enrichment of specific chromosomes for DE genes between hybrids for
267 each sample type using hypergeometric tests in R (phyper) and adjusted *P*-values to an FDR of

268 5% (Benjamini and Hochberg 1995). To reduce false positives, we used only the number of
269 autosomal DE genes as the background in the hypergeometric tests because of the known over-
270 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.

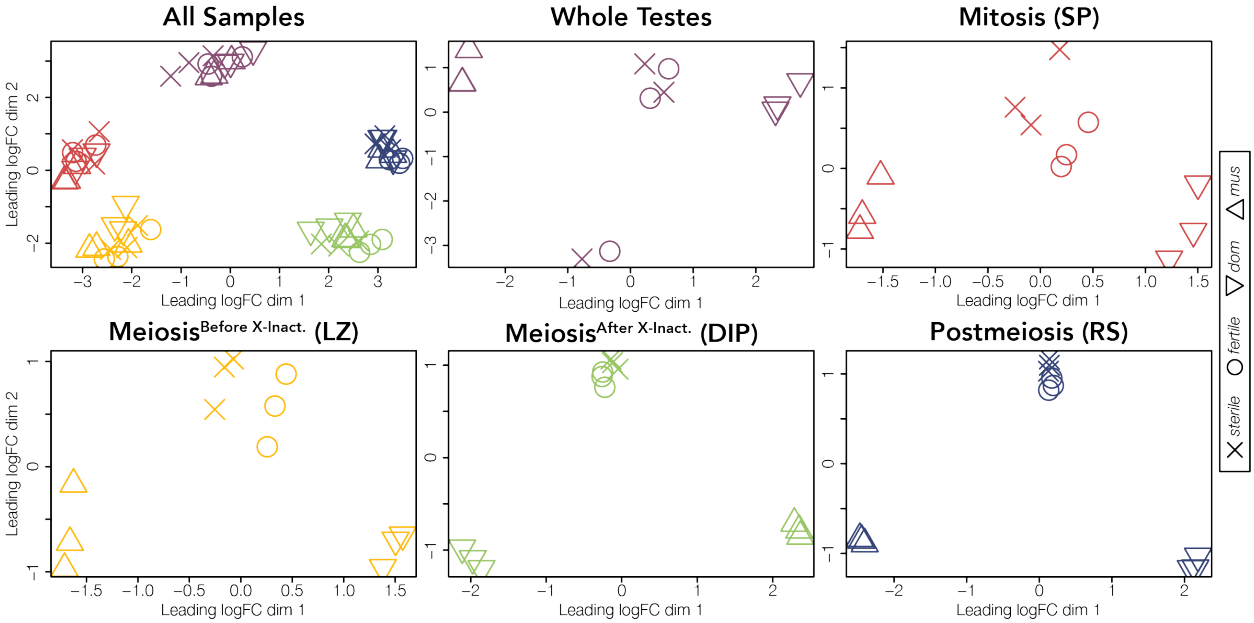


Figure 2. Sample type then cross type drives differences in expression profiles. Multidimensional scaling (MDS) plots of distances among and within sample types for expressed genes across all chromosomes. Distances are calculated as the root-mean-square deviation (Euclidean distance) of log2 fold changes 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 =

286 Postmeiosis, and purple = Whole Testes). The upper left MDS plot includes all sample types and
287 remaining plots show each sample type individually. Each sample type for each cross is
288 represented by three replicates.
289
290 Because of the apparent increased variation among whole testes hybrid samples, we next
291 quantified sample variation within both datasets. We measured variation among replicates using
292 the BCV, restricting our analysis to only protein coding genes. Whole testes had greater variation
293 among replicates (BCV = 0.347) compared to sorted cells (BCV = 0.182; Fig S2). Additionally,
294 hybrid whole testes had the greatest variation among replicates (BCV = 0.445; CI_1 =
295 [0.443,0.445]; CI_2 = [0.375,0.514]) compared to parent whole testes (BCV = 0.207; CI_1 =
296 [0.207,0.208]; CI_2 = [0.161,0.251]), parent sorted cells (BCV = 0.189; CI_1 = [0.190,0.191]; CI_2 =
297 [0.185,0.192]), and hybrid sorted cells (BCV = 0.174; CI_1 = [0.175,0.176]; CI_2 = [0.171,0.177];
298 Fig S3-S5). When including all annotated genes in variance calculations, the BCV was still
299 greater in whole testes than in sorted cell populations despite the presence of some lowly
300 expressed and highly variable non-protein coding genes in the sorted cell dataset (Figs S6, S7).
301
302 *Whole testes expression patterns are driven by diverse cell composition*
303 We next quantified expression of two panels of marker genes associated with specific
304 testes cell types in fertile reference *mus* and *dom* samples, where gene expression is not expected
305 to be disrupted. This allowed us to assess the purity of sorted cell populations as determined by
306 expression of marker genes from non-target cell types and to ascertain which cell types were
307 contributing to the unique expression patterns observed in whole testes. Our first panel included
308 marker genes associated with spermatagonia (mitosis), spermatocytes (meiosis), round

309 spermatids (postmeiosis), elongating spermatids (postmeiosis), endothelial cells, Sertoli cells
310 (support cells), and Leydig cells (testosterone producing cells), while the second panel included
311 additional cell types (from Hermann et al. 2018; Figs S8, S9). Results from both marker panels
312 were consistent. As expected, sorted cell populations mostly expressed only marker genes
313 characteristic of their target cell type, overall indicating successful FACS enrichment (results for
314 *dom* Figs 3, S8, results for *mus* Figs S9, S10). Mitotic cells showed high expression of
315 spermatogonia markers and limited expression of non-target markers indicating relative cell
316 purity. However, intermediate expression of endothelial and Sertoli markers suggested that the
317 FACS protocol for isolating this cell population may also have captured other somatic cells.
318 Meiotic$^{\text{Before X-Inact.}}$ cells appeared to have some spermatogonia contamination, while Meiotic^{After}
319 $^{\text{X-Inact.}}$ cells showed very high purity, expressing only spermatocyte-specific markers. Postmeiotic
320 cells had high expression of round spermatid markers as expected, but also some expression of
321 elongating spermatid markers indicating that FACS may also have captured the developmental
322 transition to these cells.
323

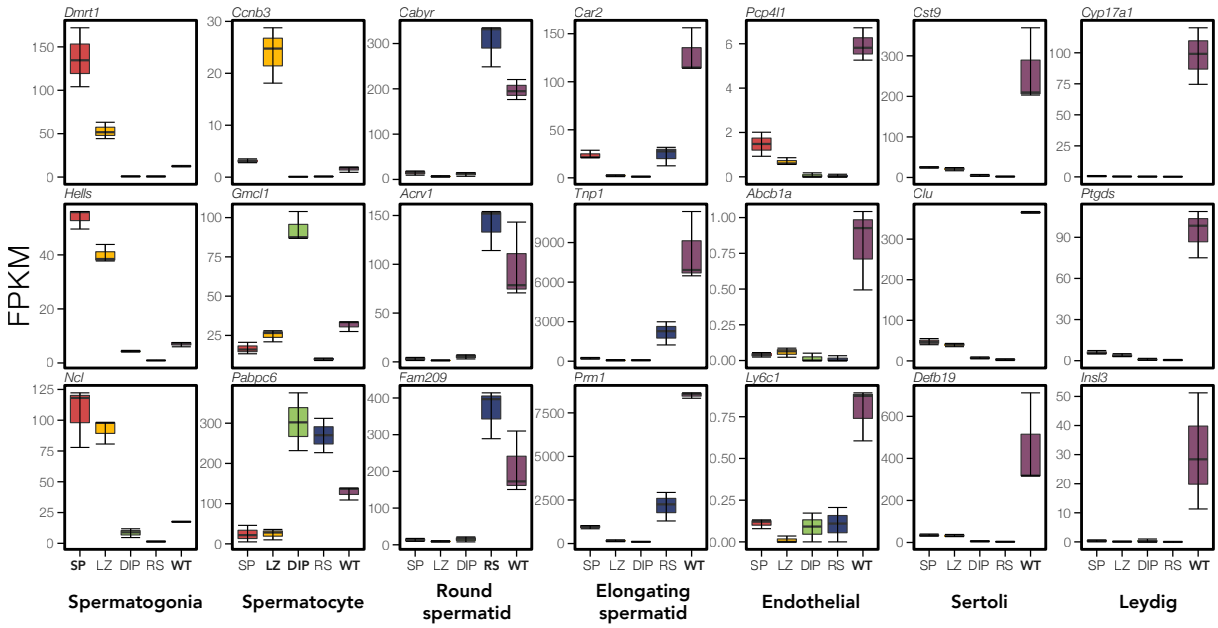
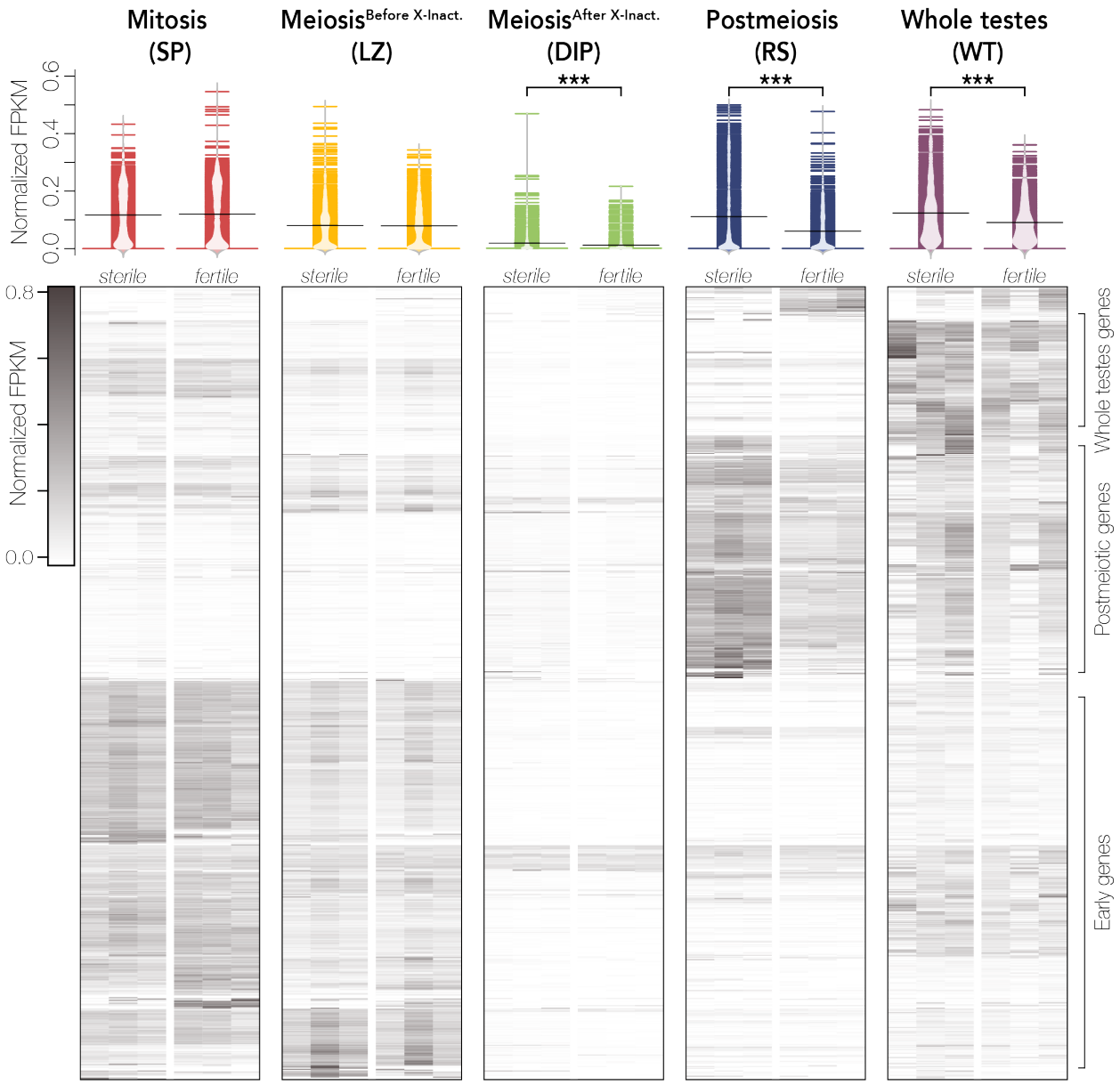


Figure 3. Whole testes show signatures of more diverse cell types than enriched cell populations. Expression of cell-specific marker genes (Green et al. 2018) across each sample type for *dom* reference samples. We quantified expression (FPKM) of three marker genes (rows) associated with testes-specific cell types (columns). Each panel displays marker expression in each sample type (red = Mitosis (SP), yellow = Meiosis$^{\text{Before X-Inact.}}$ (LZ), green = Meiosis$^{\text{After X-Inact.}}$ (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$^{\text{Before X-Inact.}}$ cells (Maekawa et al. 2004), and *Gmc1* is specific to Meiotic$^{\text{After X-Inact.}}$ cells (Nguyen et al. 2002).

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

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



347

348 **Figure 4. Patterns of X-linked gene expression in *sterile* and *fertile* hybrids differ between**

349 **sorted cells and whole testes.** The upper panel displays expression distributions (as normalized

350 FPKM) across replicates for each sample type across X-linked genes. FPKM values were

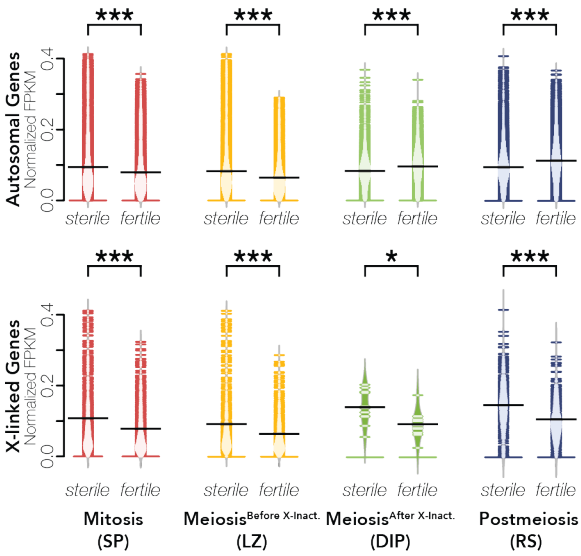
351 normalized so that the sum of squares equals one using the R package vegan (Oksanen et al.

352 2007). Expression distributions are colored by sample type (red = Mitosis, yellow = Meiosis^{Before}

353 $^{\text{X-Inact.}}$, green = Meiosis$^{\text{After X-Inact.}}$, blue = Postmeiosis, and purple = Whole Testes) and labelled by

354 cross (*sterile* or *fertile* hybrid). These plots were generated with the R package beanplot
355 (Kampstra 2008) and differences in expression were calculated with Wilcoxon signed-rank tests
356 where *** indicates $p < 0.001$ and * indicates $p < 0.05$ after FDR correction (Benjamini and
357 Hochberg 1995). The lower panel shows a heatmap of X-linked gene expression plotted as
358 normalized FPKM values that are hierarchically clustered using Euclidean distance. Each row
359 represents a gene with darker colors indicating higher expression. The heatmap was generated
360 with the R package ComplexHeatmap v.2.3.2 (Gu *et al.* 2016).
361
362 *Both changes in cellular composition of whole testes and regulatory divergence contribute to*
363 *expression differences in hybrids*
364 We further tested whether changes in cellular composition of complex tissues influences
365 relative expression levels between contrasts. Indeed, we found that differences in whole testes
366 cell composition between *sterile* and *fertile* hybrids appears to be a large driver of differences in
367 relative expression of stage-specific genes (Fig 5). In *fertile* hybrids, whole testes are largely
368 composed of late spermatogensis cell types. In *sterile* hybrids, there is a disruption in
369 development immediately before normal MSCI, which triggers an apoptotic cascade and
370 decreases downstream meiotic and postmeiotic cell abundance (Schwahn *et al.* 2018). Based on
371 these histological predictions, we expected stage-specific genes from pre-X chromosome
372 inactivation stages (Mitosis and Meiosis$^{\text{Before X-Inact.}}$) to appear over-expressed in *sterile* hybrids
373 and stage-specific genes from post-X chromosome inactivation stages (Meiosis$^{\text{After X-Inact.}}$ and
374 Postmeiosis) to appear under-expressed in *sterile* hybrids. Consistent with this, in whole testes,
375 autosomal Mitotic- and Meiotic$^{\text{Before X-Inact.}}$-specific genes had higher expression in *sterile* hybrids
376 (one-sided Wilcoxon Signed-Rank Test; autosomal Mitotic: $n = 5307$, $V = 11247685$, $p = 0$;

377 autosomal Meiotic$^{\text{Before X-Inact}}$: n = 4215, V = 7988923, p = 0), while autosomal Meiotic$^{\text{After X-Inact.}}$-
378 and Postmeiotic-specific genes had lower expression (one-sided Wilcoxon Signed-Rank Test;
379 autosomal Meiotic$^{\text{After X-Inact}}$: n = 4544, V= 2005025, p = 1.46 x 10^{-276} ; autosomal Postmeiotic: n
380 = 7417, V = 4789686, p = 0; Fig 5).



381

382 **Figure 5. Changes in cellular composition alters expression of stage-specific genes in whole**
383 **testes samples.** For each sorted cell population, we defined a set of stage-specific genes and
384 compared their expression in whole testes of *sterile* and *fertile* hybrids. Mitotic and Meiotic^{Before}
385 $^{\text{X-Inact.}}$ cells are present at lower abundances in *sterile* hybrids while Meiotic$^{\text{After X-Inact.}}$ and
386 Postmeiotic cells are present at higher abundances (Schwahn et al. 2018). FPKM is normalized
387 so that the sum of squares equals 1 using the R package vegan (Oksanen et al. 2007). Differences
388 in expression were calculated with Wilcoxon signed-rank tests where *** indicates $p < 0.001$
389 and * indicates $p < 0.05$ after FDR correction (Benjamini and Hochberg 1995).

390

391 Given the nature of hybrid sterility in house mice (Bhattacharyya et al. 2013), we had
392 different expectations for X-linked genes. The normal regulation of the X chromosome is not

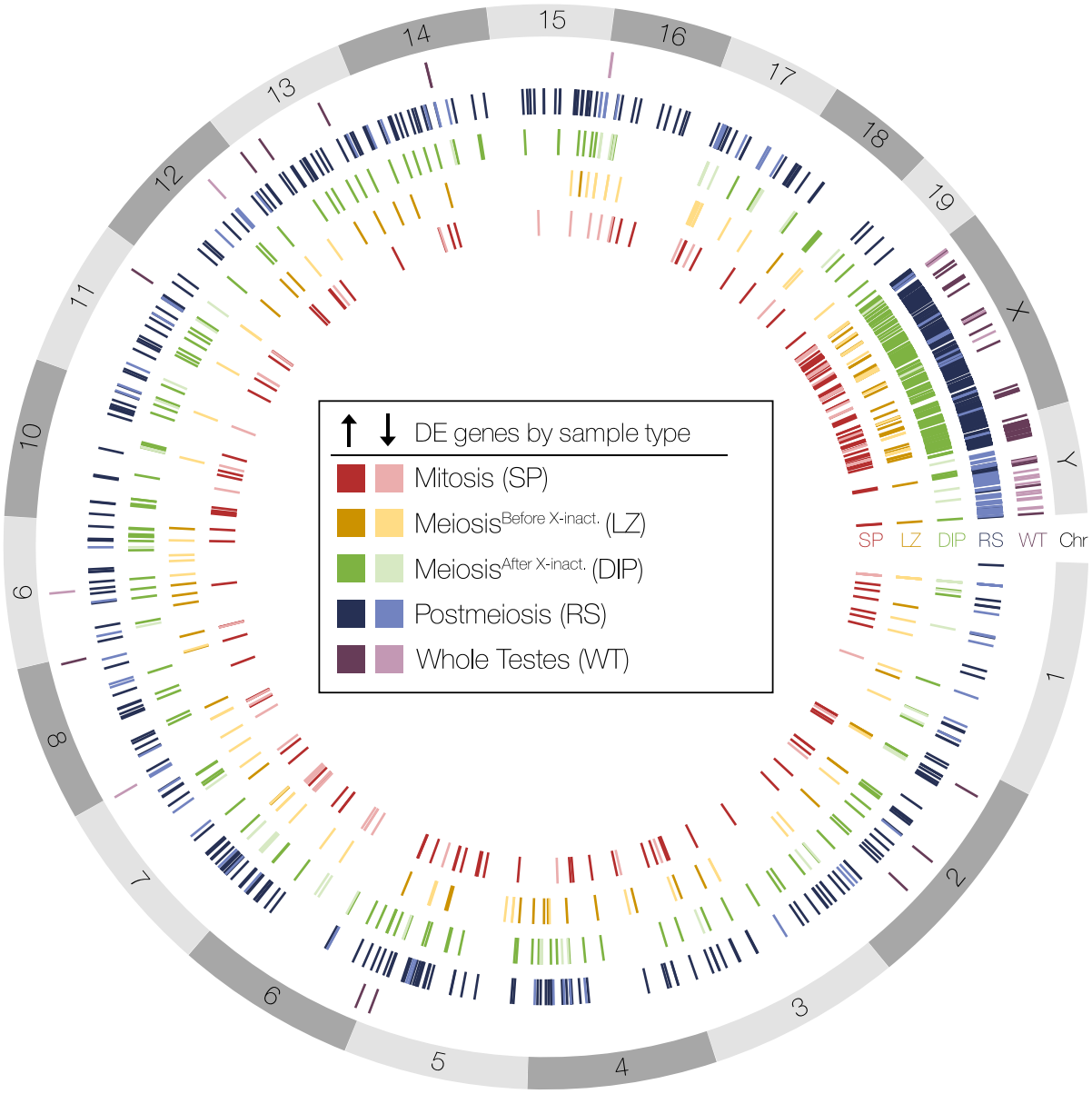
393 disrupted in pre-X inactivation cell types, so differences in cellular composition should drive
394 expression patterns for stage-specific X-linked genes in pre-X inactivation cell types as with
395 autosomal genes. However, the X chromosome is over-expressed in post-X inactivation cell
396 types (Larson et al. 2017), so changes in cellular composition and known regulatory divergence
397 could influence expression patterns of post-X inactivation stage-specific genes in *sterile* whole
398 testes. As we predicted based on cell composition, X-linked Mitotic and Meiotic$^{\text{Before X-Inact.}}$ genes
399 still had higher expression in *sterile* hybrids (one-sided Wilcoxon signed-rank test; X-linked
400 Mitotic: n = 465, V = 95946, p = 1.16 x 10^{-47}; X-linked Meiotic$^{\text{Before X-Inact}}$: n = 361, V = 60492, p
401 = 1.53 x 10^{-44}). However, X-linked Meiotic$^{\text{After X-Inact.}}$ and Postmeiotic genes also had higher
402 expression (two-sided Wilcoxon Signed-Rank Test; X-linked Meiotic$^{\text{After X-Inact.}}$: n = 11, V = 56,
403 p = 0.0420; X-linked Postmeiotic: n = 252, V = 25826, p = 1.59 x 10^{-17}), indicating that the
404 disruption of X chromosome inactivation and repression in *sterile* hybrids had a stronger effect
405 on expression patterns than changes in cell composition, despite the lower abundances of these
406 cell types (Schwahn et al. 2018). Together these results indicate that the high proportion of
407 postmeiotic cells in whole testes is a major cause of differences in expression patterns of
408 autosomal and many X-linked genes between *sterile* and *fertile* whole testes samples.
409 We further investigated the detectability of patterns of disrupted X chromosome
410 regulation in *sterile* hybrids across both sampling approaches and found that whole testes
411 sampling partially masks signatures of X chromosome misexpression. Previous research using
412 sorted cell populations has shown that disruption of MSCI in *sterile* hybrids manifests as over-
413 expression of the X chromosome both in terms of more expressed X-linked genes and higher
414 average X-linked gene expression (Larson et al. 2017). We recovered the expected pattern of
415 higher X-linked gene expression in *sterile* hybrids in both sorted cell populations (one-sided

416 Wilcoxon signed-rank test; X-linked Meiosis$^{\text{After X-Inact}}$: n = 896, V = 273584, p = 5.34 x 10^{-60}; X-
417 linked Postmeiosis: n = 896, V = 290110, p = 1.18 x 10^{-64}; Fig 4) and in whole testes (one-sided
418 Wilcoxon signed-rank test; n = 896, V = 326947, p = 2.40 x 10^{-64}; Fig 4). We also found more
419 detected X-linked genes in Meiosis$^{\text{After X-Inact.}}$ ($F_{3,8}$ = 13.8, p = 1.58 x 10^{-3}; Fig S11A) and
420 Postmeiosis ($F_{3,8}$ = 31.87, p = 8.47 x 10^{-05}; Fig S11B), but there was no difference in the number
421 of detected X-linked genes in *sterile* whole testes ($F_{3,8}$ = 0.606, p = 0.629; Fig S11C), regardless
422 of the FPKM threshold used to define detected genes.
423
424 *Whole testes sampling reduces power for differential expression inference*
425 The increased variance among replicates and the resulting decreased power in the whole
426 testes dataset also greatly reduced the number of genes considered differentially expressed
427 between *sterile* and *fertile* hybrids in whole testes compared with sorted cell populations (Fig 6;
428 Table S2; Supplemental File 2). Fewer DE genes were detected between hybrids for whole testes
429 samples (DE genes = 83; Table S2) compared to sorted cell populations (Mitotic DE genes =
430 231, Meiotic$^{\text{Before X-Inact.}}$ DE genes = 178, Meiotic$^{\text{After X-Inact.}}$ DE genes = 343, and Postmeiotic DE
431 genes = 606). However, both whole testes and sorted cell populations exhibited similar broad
432 patterns of differential expression. In both datasets, more DE genes were upregulated in *sterile*
433 hybrids than were downregulated (Table S2), and we were able to detect enrichment of the X and
434 Y chromosomes for DE genes as previously reported (Larson et al. 2017; Fig S12; Tables S3,
435 S4). In addition, no DE genes between *sterile* and *fertile* hybrids were differentially up- or down-
436 regulated in whole testes samples compared to sorted cell populations (Table S5; Fig S13),
437 although we did find this pattern when comparing DE genes between *mus* and *dom* mice— a

438 small proportion of genes were differentially regulated in whole testes samples compared to

439 sorted cell populations (0.43% - 3.16%; Table S6; Supplemental Files S3, S4).

440



441

442 **Figure 6. Whole testes and enriched cell populations differed in the number and identity of**

443 **dfferentially expressed genes.** Spatial distribution of differentially expressed (DE) genes across

444 reference mouse genome chromosomes (build GRCm38.p6) between *sterile* and *fertile* hybrids

445 for all five sample types. Darker colors indicate genes up-regulated in *sterile* hybrids and lighter
446 colors indicated genes down-regulated in *sterile* hybrids.
447
448 Despite consistent patterns of enrichment of DE genes on the sex chromosomes and
449 direction of expression of DE genes between hybrids, there was very little overlap in DE genes
450 between each sample type. Whole testes samples shared very few genes in common with any of
451 the sorted cell populations (Fig 7). Additionally, there were very few DE genes shared across the
452 different stages of spermatogenesis, although the proportion of DE genes shared between sample
453 types generally increased with stricter fold change cutoffs (Table S7). Sorted cell samples often
454 have large repertoires of genes that were only differentially expressed within one cell type (Fig
455 7) though there was greater overlap of DE genes between post-X inactivation cell types
456 (Meiosis$^{\text{After X-Inact.}}$ and Postmeiosis). In sum, different sampling methodology clearly altered the
457 overall and gene-specific resolution of the regulatory underpinnings of hybrid male sterility.
458
459

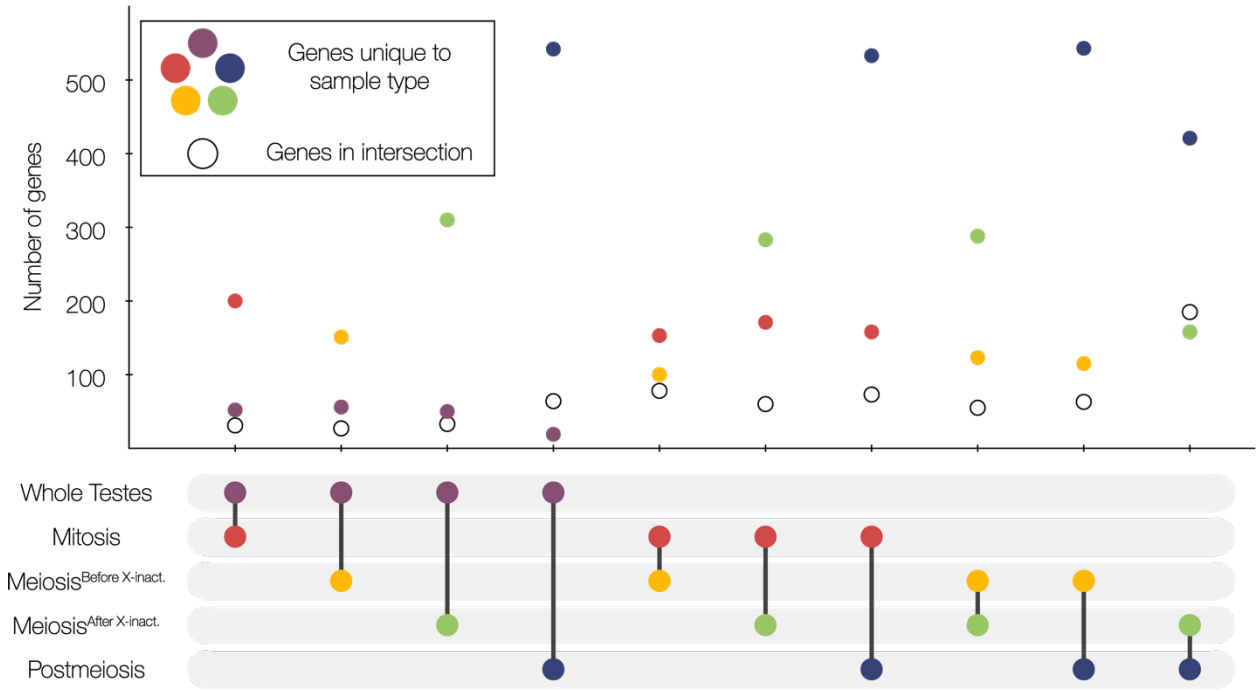


Figure 7. Whole testes and enriched cell populations differed in pairwise comparisons of DE genes between *sterile* and *fertile* hybrids. The sample types in each comparison are indicated by the pair of connected dots in the bottom panel. For each comparison, DE genes common between the two sample types 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$^{\text{Before X-Inact.}}$, green = Meiosis$^{\text{After X-Inact.}}$, blue = Postmeiosis, and purple = Whole Testes).

475 DISCUSSION

476 *Transcriptomic biases of complex tissues in evolutionary biology*

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

497 Reproductive tissues are likely to be particularly prone to structural divergence, as
498 cellular composition is expected to evolve in response to selection for increased reproductive
499 success. For example, sperm competition leads to selection for males to increase sperm numbers
500 (Firman et al. 2013, 2018) or the proportion of sperm-producing tissue within the testes (Lüpold
501 et al. 2009). Sperm production can be increased in multiple ways, each of which has different
502 consequences for the cellular architecture of the testis (Schärer *et al.* 2011; Ramm and Schärer
503 2014). The non-sperm-producing tissue within the testes can also evolve in response to sexual
504 selection. An extreme example are Capybara, which devote ~30% of their testes to the
505 testosterone-producing Leydig cells (in sharp contrast to other rodents, where Leydig cells
506 comprise only 2.7-5.3% of testes; Costa et al. 2006; Lara et al. 2018). Differences in
507 reproductive investment can also drive apparent expression differences between species. Gene
508 expression divergence between humans and chimpanzees is elevated in testes relative to other
509 tissues, a pattern proposed to reflect positive selection on gene expression levels (Khaitovich et
510 al. 2005, 2006). However, whole testis transcriptomes tend to be more similar between species
511 with similar mating systems and cellular architectures (Brawand et al. 2011; Yapar et al. 2021),
512 which have presumably evolved convergently in response to investment in sperm production.
513 Our results show that in bulk tissues even minor testis cell types (such as Leydig cells and Sertoli
514 cells) contribute to overall expression profiles and suggest that differences in the proportion of
515 any cell type have the potential to strongly modify expression profiles of whole tissues.
516
517 *Reducing sample complexity in evolutionary studies of expression divergence*
518 Here we confirm that FACS is an effective way of isolating relatively pure cell types and
519 removing the effects of divergent cellular composition from experimental contrasts (Getun et al.

520 2011; da Cruz et al. 2016; Larson et al. 2016, 2017; Geisinger et al. 2021; Kopania et al. 2021).
521 There are of course, alternative methods available for bulk cell enrichment, such as gradient
522 centrifugation to separate testes cell types (Shima et al. 2004; Chalmel et al. 2007; Rolland et al.
523 2009). These approaches are well suited for testes, given the dramatic changes in cell size, DNA
524 content, and chromatin condensation during spermatogenesis (Bellvé 1993; Getun et al. 2011),
525 but mechanical or flow cytometry-based enrichment has also been developed for other complex
526 heterogeneous tissues (e.g., late term placenta; Li et al. 2020). There is the potential for FACS to
527 bias gene expression in enriched cell populations (*i.e.,* by triggering a stress response from
528 nozzle pressure or UV exposure; Box et al. 2020). However, cell sorting procedures appear to
529 have a minimal effect on overall expression profiles, and altered expression is likely consistent
530 across treatments within an experiment and can be mitigated by minimizing the time between
531 cell sorting, RNA extraction, and storage (Box et al. 2020). Beyond the limited and potentially
532 tissue-specific methods of bulk cell enrichment, recent advances in single cell sequencing
533 technology (scRNA-Seq) and spatial transcriptomic methods (*e.g.,* sci-SPACE, Srivatsan et al.
534 2021) can allow researchers to assay a greater number of cell types across many tissue types
535 without *a priori* identification or labelling (Kiselev et al. 2019). Although both FACS and
536 scRNA-Seq are both powerful approaches for studying gene expression evolution in tissues with
537 cellular composition differences (Kopania et al. 2021; Murat et al. 2021), they are both currently
538 difficult to apply in non-model systems, especially for field-based studies, as they typically
539 require access to flow cytometers and a short timeline for tissue biopsy, cell sorting, and RNA
540 extraction (Getun et al. 2011; Bageritz and Raddi 2019, but see also Wohnhaas et al. 2019;
541 Denisenko et al. 2020). Additionally, scRNA-Seq protocols likely have some of the same sources

542 of biased gene expression as FACS, and further research should be done to determine how
543 different enrichment protocols alter expression inferences.

544 When cell enrichment protocols are not feasible, alternative methods are available for
545 minimizing developmental or cellular complexity differences between species or experimental
546 contrasts. For example, different stages of sperm development can be isolated by sampling whole
547 testes at different points in early sexual development (Schultz et al. 2003; Shima et al. 2004;
548 Laiho et al. 2013), across annual reproductive cycles (Rolland et al. 2009), or spatially, as in
549 *Drosophila*, where sperm develop in tubular testes, allowing dissection of distinct regions that
550 are enriched for particular cell types (Meiklejohn et al. 2011; Landeen et al. 2016). Furthermore,
551 some developmentally heterogenous samples can be artificially synchronized, for example by
552 shaving hair or plucking feathers and sampling across regrowth timelines (Poelstra et al. 2014,
553 2015; Ferreira et al. 2017). Microdissection of complex tissues is also a feasible way to minimize
554 the effects of cellular composition on transcriptomic profiles. For example, laser capture
555 microdissection provides a means to rapidly and precisely isolate cellular populations from
556 complex tissues (Emmert-Buck et al. 1996), albeit with the added requirement of highly
557 specialized instrumentation. It is common in behavioral research to dissect out major regions of
558 the brain rather than sampling the whole brain (Khrameeva et al. 2020; Sato et al. 2020). Thus, a
559 chemical or mechanical approach to partitioning complex tissues can provide researchers with a
560 way of minimizing the negative effects associated with bulk RNASeq in their own studies.

561 Despite the potentially confounding effects of cellular composition and regulatory
562 divergence in whole tissue sampling, a bulk RNASeq approach is appropriate in cases where a
563 cell type of interest is not easily isolated or when researchers wish to capture all developmental
564 stages. For example, Larson et al. (2017) used FACS to isolate only four stages of

565 spermatogenesis, but postzygotic isolation barriers can operate at many different stages of
566 spermatogenesis (Oka et al. 2010; Ishishita et al. 2015; Torgasheva and Borodin 2016; Schwahn
567 et al. 2018; Yoshikawa et al. 2018; Liang and Sharakhov 2019). In these situations, bulk
568 RNASeq can allow researchers to investigate expression differences in hard to obtain cell types.
569 Indeed, some evolutionary inferences may be robust to sampling strategy. The misexpressed
570 genes in hybrids identified by Mack et al. (2016) overlapped substantially with sterility eQTLs
571 identified in wild hybrids from natural hybrid zones of *M. m. musculus* and *M. m. domesticus*
572 populations (Turner and Harr 2014), suggesting that despite the decreased power and
573 susceptibility to artifacts introduced by differences in cellular composition associated with bulk
574 tissue sampling, the genes that are identified are likely genes of large effect and have a high
575 likelihood of being biologically meaningful. For all these reasons, bulk tissue sampling may be
576 an appropriate first step depending on the system and questions being addressed.

577 It is also possible to use computational approaches, such as *in silico* deconvolution
578 methods to estimate changes in cell type proportions across samples or quantify cell type-
579 specific expression profiles (Shen-Orr and Gaujoux 2013; Avila Cobos et al. 2018; Newman et
580 al. 2019). These methods rely on expression profiles from single-cell data and accurate estimates
581 of cellular proportions (Shen-Orr and Gaujoux 2013; Avila Cobos et al. 2018), which can be
582 challenging to obtain in non-model systems but are likely to become increasingly more
583 accessible as technologies advance. Deconvolution may also be less accurate when the
584 expression of specific genes varies across stages because the net expression of a gene in a whole
585 tissue may differ from its stage-specific expression. While we found that DE genes between
586 *sterile* and *fertile* hybrids had consistent direction of differential expression between our whole
587 testes samples and sorted cell populations, in our comparisons of DE genes between *mus* and

588 *dom* mice, we found DE genes that had the opposite regulation patterns between sample types.
589 Deconvolution methods in studies of hybrid misexpression may also be inherently flawed given
590 that there is often no single "sterile" phenotype (Good et al. 2008; Turner et al. 2012; Larson et
591 al. 2017; Bikchurina et al. 2018) and that the reference expression profiles used for
592 deconvolution may be disrupted in hybrids (Landeen et al. 2016; Morgan et al. 2020; Mugal et
593 al. 2020; Brekke et al. 2021). Given these drawbacks, we advocate that detailed histological
594 analysis of how the phenotype of interest manifests in complex, heterogenous tissues (Oka et al.
595 2010; Schwahn et al. 2018) should accompany any evolutionary study based on comparative
596 transcriptomic data, so that researchers can mediate biases associated with sampling
597 methodology when designing future studies.
598
599 *Power to detect differential expression using bulk RNASeq*
600 The primary analytical goal of most RNASeq studies is to identify DE genes. It is vital
601 that we can accurately determine which genes are differentially expressed because we use these
602 patterns for a myriad of downstream analyses. Accurate assessment should also increase
603 resolution into the genomic basis of phenotypes of interest. We found that bulk RNASeq can
604 hinder differential expression analyses through an increase in replicate variability, potentially
605 masking biologically meaningful changes in gene expression. RNASeq analyses are sensitive to
606 both technical and biological variation (Todd et al. 2016), and studies of outbred wild
607 populations are inherently disadvantaged because of the power lost from increased biological
608 variation (Liu et al. 2014; Todd et al. 2016). The BCV is an estimate of the variation among
609 biological replicates and is correlated with power to detect DE genes. We found that in inbred
610 strains of house mice, whole testes had higher inter-replicate variability in expression than sorted

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

624 Ultimately, both whole tissue and cell enrichment-based approaches were able to detect
625 broad-scale patterns of disrupted sex chromosome expression in *sterile* hybrids. In house mice,
626 MSCI is disrupted in *sterile* hybrids (Bhattacharyya et al. 2013; Davies et al. 2016; Gregorova et
627 al. 2018), leading to an over-expression of X-linked genes (Good et al. 2010; Campbell et al.
628 2013; Turner et al. 2014). Both Mack *et al.* (2016) and Larson *et al.* (2017) found higher
629 expression of genes across the X chromosome in *sterile* hybrids, but our results show that it is
630 more difficult to detect an increased number of expressed X-linked DE genes between *sterile*
631 hybrids and their parents using whole testes sampling. Patterns of X-linked over-expression can
632 also be recovered in whole testes given *a priori* knowledge of stage-specific genes for cell types
633 where the X chromosome should be inactivated or repressed. Of course, approaches relying on

634 orthologous sets of stage-specific genes from other species will be limited to species with close
635 evolutionary relationships to model organisms. A sensitivity of the regulatory mechanisms
636 controlling sex chromosome expression during male meiosis has been proposed to be a major
637 mechanism underlying hybrid sterility (Lifschytz and Lindsley 1972), but so far, genomic
638 evidence for disrupted MSCI and downstream postmeiotic repression in other mammalian taxa is
639 conflicting. In sterile hybrid cats, the X chromosome is misexpressed and MSCI is disrupted
640 (Davis et al. 2015; Bredemeyer et al. 2021), while sterile rabbit hybrids do not support a role of
641 X chromosome misexpression in speciation (Rafati *et al.* 2018). Studies outside of house mice
642 have largely relied on bulk whole testes sequencing (but see Bredemeyer et al. 2021) and
643 understanding if the detected or undetected misexpression of the X is biologically accurate is
644 important for determining the role of disrupted sex chromosome regulation in postzygotic
645 isolation and speciation. Using targeted approaches can give us the developmental perspective
646 needed for contextualizing the origins of reproductive barriers (Cutter and Bundus 2020).
647

648 *Conclusions*

649 Here, we demonstrate important consequences of differing cell composition in
650 identifying DE genes in the context of hybrid sterility. We advocate for sampling approaches
651 which allow for developmental perspectives in RNASeq studies, so that we can accurately probe
652 species barriers. These same issues are important for other evolutionary contrasts in complex
653 tissues, and we underscore the importance of considering the cellular and developmental context
654 of complex expression in evolutionary studies. Our results suggest that sampling methodology
655 could influence the biological implications of not only hybrid misexpression in speciation, but
656 also across studies of divergent gene expression broadly. The consequences of whole tissue

657 sampling of complex tissues have the potential to alter not only inferred gene ontological
658 processes, but also the structure and evolution of gene networks, the relative importance of cis-
659 and trans-regulatory evolution, and even insights into the processes and rates underlying
660 expression evolution.
661
662 **Acknowledgements:** This work was supported by an NSF Graduate Research Fellowship to
663 KEH (DGE- 2034612), an NSF grant to ELL (DEB 1557059), and grants from the Eunice
664 Kennedy Shriver National Institute of Child Health and Human Development of the National
665 Institutes of Health (R01-HD073439, R01-HD094787) to JMG. We would like to thank Jonathan
666 Velotta and members of the Larson and Tinghitella Labs for feedback on this project and Ben
667 Fotovich, Ivan Kovanda, and the IT Department at the University of Denver for support using
668 the High Performance Computing Cluster.
669
670 **Author contributions:** KEH, ELL, and JMG conceived of the study. KEH conducted the
671 analyses. KEH and ELL wrote the manuscript with input from JMG.
672
673 **Data accessibility**: There is no data to be archived. Scripts used in the manuscript will be
674 available upon publication at https://github.com/KelsieHunnicutt.
675
676 **Conflict of interest:** The authors declare no conflict of interest.
677
678
679

680 LITERATURE CITED

681 Abolins-Abols, M., E. Kornobis, P. Ribeca, K. Wakamatsu, M. P. Peterson, E. D. Ketterson, and
682 B. Milá. 2018. Differential gene regulation underlies variation in melanic plumage
683 coloration in the dark-eyed junco (Junco hyemalis). Mol. Ecol. 27:4501–4515.
684 Alvarez, M., A. W. Schrey, and C. L. Richards. 2015. Ten years of transcriptomics in wild
685 populations: what have we learned about their ecology and evolution? Mol. Ecol.
686 24:710–725.
687 Anders, S., and W. Huber. 2010. Differential expression analysis for sequence count data. Nat.
688 Prec., doi: 10.1038/npre.2010.4282.2.
689 Avila Cobos, F., J. Vandesompele, P. Mestdagh, and K. De Preter. 2018. Computational
690 deconvolution of transcriptomics data from mixed cell populations. Bioinformatics
691 34:1969–1979.
692 Bageritz, J., and G. Raddi. 2019. Single-cell RNA sequencing with Drop-Seq. Methods Mol.
693 Biol. 1979:73–85.
694 Bellvé, A. R. 1993. Purification, culture, and fractionation of spermatogenic cells. Pp. 84–113 *in*
695 P. M. Wassarman and M. L. DePamphilis, eds. Methods in Enzymology. Academic
696 Press, London.
697 Bellvé, A. R., J. C. Cavicchia, C. F. Millette, D. A. O'Brien, Y. M. Bhatnagar, and M. Dym.
698 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological
699 characterization. J. Cell Biol. 74:68–85.
700 Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and
701 powerful approach to multiple testing. J. R. Stat. Soc. 57:289–300.

702 Bhattacharyya, T., S. Gregorova, O. Mihola, M. Anger, J. Sebestova, P. Denny, P. Simecek, and
703 J. Forejt. 2013. Mechanistic basis of infertility of mouse intersubspecific hybrids. Proc.
704 Natl. Acad. Sci. U. S. A. 110:E468–E477.
705 Bikchurina, T. I., K. V. Tishakova, E. A. Kizilova, S. A. Romanenko, N. A. Serdyukova, A. A.
706 Torgasheva, and P. M. Borodin. 2018. Chromosome synapsis and recombination in male-
707 sterile and female-fertile interspecies hybrids of the dwarf hamsters (Phodopus,
708 Cricetidae). Genes 9:227.
709 Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina
710 sequence data. Bioinformatics 30:2114–2120.
711 Box, A., M. DeLay, S. Tighe, S. V. Chittur, A. Bergeron, M. Cochran, P. Lopez, E. M. Meyer,
712 A. Saluk, S. Thornton, and K. Brundage. 2020. Evaluating the Effects of Cell Sorting on
713 Gene Expression. J. Biomol. Tech. 31:100–111.
714 Brawand, D., M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, P. Harrigan, M. Weier, A.
715 Liechti, A. Aximu-Petri, M. Kircher, F. W. Albert, U. Zeller, P. Khaitovich, F. Grützner,
716 S. Bergmann, R. Nielsen, S. Pääbo, and H. Kaessmann. 2011. The evolution of gene
717 expression levels in mammalian organs. Nature 478:343–348.
718 Bredemeyer, K. R., C. M. Seabury, M. J. Stickney, J. R. McCarrey, B. M. vonHoldt, and W. J.
719 Murphy. 2021. Rapid macrosatellite evolution promotes X-Linked hybrid male sterility in
720 a feline interspecies cross. Mol. Biol. Evol., doi: 10.1093/molbev/msab274.
721 Brekke, T. D., E. C. Moore, S. C. Campbell-Staton, C. M. Callahan, Z. A. Cheviron, and J. M.
722 Good. 2021. X chromosome-dependent disruption of placental regulatory networks in
723 hybrid dwarf hamsters. Genetics 218:iyab043.

724 Breschi, A., T. R. Gingeras, and R. Guigó. 2017. Comparative transcriptomics in human and
725 mouse. Nat. Rev. Genet. 18:425–440.
726 Britton-Davidian, J., F. Fel-Clair, J. Lopez, P. Alibert, and P. Boursot. 2005. Postzygotic
727 isolation between the two European subspecies of the house mouse: estimates from
728 fertility patterns in wild and laboratory-bred hybrids. Biol. J. Linn. Soc. Lond. 84:379–
729 393.
730 Buchberger, E., M. Reis, T.-H. Lu, and N. Posnien. 2019. Cloudy with a chance of insights:
731 context dependent gene regulation and implications for evolutionary studies. Genes
732 10:492.
733 Campbell, P., J. M. Good, and M. W. Nachman. 2013. Meiotic sex chromosome inactivation is
734 disrupted in sterile hybrid male house mice. Genetics 193:819–828.
735 Carlson, B. A., S. M. Hasan, M. Hollmann, D. B. Miller, L. J. Harmon, and M. E. Arnegard.
736 2011. Brain evolution triggers increased diversification of electric fishes. Science
737 332:583–586.
738 Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of
739 morphological evolution. Cell 134:25–36.
740 Catron, D. J., and M. A. F. Noor. 2008. Gene expression disruptions of organism versus organ in
741 Drosophila species hybrids. PLoS One 3:e3009.
742 Chalmel, F., A. D. Rolland, C. Niederhauser-Wiederkehr, S. S. W. Chung, P. Demougin, A.
743 Gattiker, J. Moore, J.-J. Patard, D. J. Wolgemuth, B. Jégou, and M. Primig. 2007. The
744 conserved transcriptome in human and rodent male gametogenesis. Proc. Natl. Acad. Sci.
745 U. S. A. 104:8346–8351.

746 Costa, D. S., T. A. R. Paula, and S. L. P. Matta. 2006. The intertubular compartment
747 morphometry in capybaras (Hydrochoerus hydrochaeris) testis. Anim. Reprod. Sci.
748 91:173–179.
749 Cutter, A. D., and J. D. Bundus. 2020. Speciation and the developmental alarm clock. Elife
750 9:e56276.
751 da Cruz, I., R. Rodríguez-Casuriaga, F. F. Santiñaque, J. Farías, G. Curti, C. A. Capoano, G. A.
752 Folle, R. Benavente, J. R. Sotelo-Silveira, and A. Geisinger. 2016. Transcriptome
753 analysis of highly purified mouse spermatogenic cell populations: gene expression
754 signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. BMC
755 Genomics 17:294.
756 Davidson, E. H., and D. H. Erwin. 2006. Gene regulatory networks and the evolution of animal
757 body plans. Science 311:796–800.
758 Davidson, J. H., and C. N. Balakrishnan. 2016. Gene regulatory evolution during speciation in a
759 songbird. G3 6:1357–1364.
760 Davies, B., E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli,
761 D. Biggs, R. Diaz, C. Preece, R. Li, E. Bitoun, K. Brick, C. M. Green, R. D. Camerini-
762 Otero, S. R. Myers, and P. Donnelly. 2016. Re-engineering the zinc fingers of PRDM9
763 reverses hybrid sterility in mice. Nature 530:171–176.
764 Davis, B. W., C. M. Seabury, W. A. Brashear, G. Li, M. Roelke-Parker, and W. J. Murphy.
765 2015. Mechanisms underlying mammalian hybrid sterility in two feline interspecies
766 models. Mol. Biol. Evol. 32:2534–2546.
767 Denisenko, E., B. B. Guo, M. Jones, R. Hou, L. de Kock, T. Lassmann, D. Poppe, O. Clément,
768 R. K. Simmons, R. Lister, and A. R. R. Forrest. 2020. Systematic assessment of tissue

769 dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows.
770 Genome Biol. 21:130.
771 Emmert-Buck, M. R., R. F. Bonner, P. D. Smith, R. F. Chuaqui, Z. Zhuang, S. R. Goldstein, R.
772 A. Weiss, and L. A. Liotta. 1996. Laser capture microdissection. Science 274:998–1001.
773 Ernst, C., N. Eling, C. P. Martinez-Jimenez, J. C. Marioni, and D. T. Odom. 2019. Staged
774 developmental mapping and X chromosome transcriptional dynamics during mouse
775 spermatogenesis. Nat. Commun. 10:1251.
776 Ferreira, M. S., P. C. Alves, C. M. Callahan, J. P. Marques, L. S. Mills, J. M. Good, and J. Melo-
777 Ferreira. 2017. The transcriptional landscape of seasonal coat colour moult in the
778 snowshoe hare. Mol. Ecol. 26:4173–4185.
779 Firman, R. C., F. Garcia-Gonzalez, L. W. Simmons, and G. I. André. 2018. A competitive
780 environment influences sperm production, but not testes tissue composition, in house
781 mice. J. Evol. Biol. 31:1647–1654.
782 Firman, R. C., F. Garcia-Gonzalez, E. Thyer, S. Wheeler, Z. Yamin, M. Yuan, and L. W.
783 Simmons. 2015. Evolutionary change in testes tissue composition among experimental
784 populations of house mice. Evolution 69:848–855.
785 Firman, R. C., I. Klemme, and L. W. Simmons. 2013. Strategic adjustments in sperm production
786 within and between two island populations of house mice. Evolution 67:3061–3070.
787 Geisinger, A., R. Rodríguez-Casuriaga, and R. Benavente. 2021. Transcriptomics of meiosis in
788 the male mouse. Front Cell Dev Biol 9:446.
789 Getun, I. V., B. Torres, and P. R. J. Bois. 2011. Flow cytometry purification of mouse meiotic
790 cells. J. Vis. Exp. e2602.

791 Good, J. M., T. Giger, M. D. Dean, and M. W. Nachman. 2010. Widespread over-expression of
792 the X chromosome in sterile F1 hybrid mice. PLoS Genet. 6:e1001148.
793 Good, J. M., M. A. Handel, and M. W. Nachman. 2008. Asymmetry and polymorphism of
794 hybrid male sterility during the early stages of speciation in house mice. Evolution
795 62:50–65.
796 Good, J. M., and M. W. Nachman. 2005. Rates of protein evolution are positively correlated with
797 developmental timing of expression during mouse spermatogenesis. Mol. Biol. Evol.
798 22:1044–1052.
799 Green, C. D., Q. Ma, G. L. Manske, A. N. Shami, X. Zheng, S. Marini, L. Moritz, C. Sultan, S. J.
800 Gurczynski, B. B. Moore, M. D. Tallquist, J. Z. Li, and S. S. Hammoud. 2018. A
801 comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-Seq.
802 Dev. Cell 46:651-667.e10.
803 Gregorova, S., V. Gergelits, I. Chvatalova, T. Bhattacharyya, B. Valiskova, V. Fotopulosova, P.
804 Jansa, D. Wiatrowska, and J. Forejt. 2018. Modulation of Prdm9-controlled meiotic
805 chromosome asynapsis overrides hybrid sterility in mice. Elife 7:e34282.
806 Gu, Z., R. Eils, and M. Schlesner. 2016. Complex heatmaps reveal patterns and correlations in
807 multidimensional genomic data. Bioinformatics 32:2847–2849.
808 Handel, M. A. 2004. The XY body: a specialized meiotic chromatin domain. Exp. Cell Res.
809 296:57–63.
810 Hermann, B. P., K. Cheng, A. Singh, L. Roa-De La Cruz, K. N. Mutoji, I.-C. Chen, H.
811 Gildersleeve, J. D. Lehle, M. Mayo, B. Westernströer, N. C. Law, M. J. Oatley, E. K.
812 Velte, B. A. Niedenberger, D. Fritze, S. Silber, C. B. Geyer, J. M. Oatley, and J. R.

813 McCarrey. 2018. The mammalian spermatogenesis single-cell transcriptome, from
814 spermatogonial stem cells to spermatids. Cell Rep. 25:1650-1667.e8.
815 Holt, J., S. Huang, L. McMillan, and W. Wang. 2013. Read annotation pipeline for high-
816 throughput sequencing data. Pp. 605–612 *in* Proceedings of the International Conference
817 on Bioinformatics, Computational Biology and Biomedical Informatics - BCB'13. ACM
818 Press, New York, New York, USA.
819 Huang, S., J. Holt, C.-Y. Kao, L. McMillan, and W. Wang. 2014. A novel multi-alignment
820 pipeline for high-throughput sequencing data. Database 2014:bau057.
821 Huang, S., C.-Y. Kao, L. McMillan, and W. Wang. 2007. Transforming genomes using MOD
822 files with applications. Pp. 595–604 *in* Proceedings of the International Conference on
823 Bioinformatics, Computational Biology and Biomedical Informatics - BCB'13. ACM
824 Press, New York, New York, USA.
825 Ishishita, S., K. Tsuboi, N. Ohishi, K. Tsuchiya, and Y. Matsuda. 2015. Abnormal pairing of X
826 and Y sex chromosomes during meiosis I in interspecific hybrids of Phodopus campbelli
827 and P. sungorus. Sci. Rep. 5:9435.
828 Kampstra, P. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. J.
829 Stat. Softw. 28.
830 Khaitovich, P., I. Hellmann, W. Enard, K. Nowick, M. Leinweber, H. Franz, G. Weiss, M.
831 Lachmann, and S. Pääbo. 2005. Parallel patterns of evolution in the genomes and
832 transcriptomes of humans and chimpanzees. Science 309:1850–1854.
833 Khaitovich, P., J. Kelso, H. Franz, J. Visagie, T. Giger, S. Joerchel, E. Petzold, R. E. Green, M.
834 Lachmann, and S. Pääbo. 2006. Functionality of intergenic transcription: an evolutionary
835 comparison. PLoS Genet. 2:e171.

836 Khrameeva, E., I. Kurochkin, D. Han, P. Guijarro, S. Kanton, M. Santel, Z. Qian, S. Rong, P.
837 Mazin, M. Sabirov, M. Bulat, O. Efimova, A. Tkachev, S. Guo, C. C. Sherwood, J. G.
838 Camp, S. Pääbo, B. Treutlein, and P. Khaitovich. 2020. Single-cell-resolution
839 transcriptome map of human, chimpanzee, bonobo, and macaque brains. Genome Res.
840 30:776–789.
841 King, M. C., and A. C. Wilson. 1975. Evolution at two levels in humans and chimpanzees.
842 Science 188:107–116.
843 Kiselev, V. Y., T. S. Andrews, and M. Hemberg. 2019. Challenges in unsupervised clustering of
844 single-cell RNA-seq data. Nat. Rev. Genet. 20:273–282.
845 Kopania, E. E. K., E. L. Larson, C. Callahan, S. Keeble, and J. M. Good. 2021. Molecular
846 evolution across mouse spermatogenesis. bioRxiv, doi: 10.1101/2021.08.04.455131.
847 Kousathanas, A., D. L. Halligan, and P. D. Keightley. 2014. Faster-X adaptive protein evolution
848 in house mice. Genetics 196:1131–1143.
849 Laiho, A., N. Kotaja, A. Gyenesei, and A. Sironen. 2013. Transcriptome profiling of the murine
850 testis during the first wave of spermatogenesis. PLoS One 8:e61558.
851 Landeen, E. L., C. A. Muirhead, L. Wright, C. D. Meiklejohn, and D. C. Presgraves. 2016. Sex
852 chromosome-wide transcriptional suppression and compensatory cis-regulatory evolution
853 mediate gene expression in the Drosophila male germline. PLoS Biol. 14:e1002499.
854 Lara, N. L. M., G. M. J. Costa, G. F. Avelar, S. M. S. N. Lacerda, R. A. Hess, and L. R. de
855 França. 2018. Testis physiology—overview and histology. Pp. 105–116 *in* M. K. Skinner,
856 ed. Encyclopedia of Reproduction. Academic Press, London.
857 Larson, E. L., S. Keeble, D. Vanderpool, M. D. Dean, and J. M. Good. 2017. The composite
858 regulatory basis of the large X-effect in mouse speciation. Mol. Biol. Evol. 34:282–295.

859 Larson, E. L., E. E. K. Kopania, and J. M. Good. 2018. Spermatogenesis and the evolution of
860 mammalian sex chromosomes. Trends Genet. 34:722–732.
861 Larson, E. L., D. Vanderpool, S. Keeble, M. Zhou, B. A. J. Sarver, A. D. Smith, M. D. Dean, and
862 J. M. Good. 2016. Contrasting levels of molecular evolution on the mouse X
863 chromosome. Genetics 203:1841–1857.
864 Li, H., Q. Huang, Y. Liu, and L. X. Garmire. 2020. Single cell transcriptome research in human
865 placenta. Reproduction 160:R155–R167.
866 Li, Y.-F., W. He, K. N. Jha, K. Klotz, Y.-H. Kim, A. Mandal, S. Pulido, L. Digilio, C. J.
867 Flickinger, and J. C. Herr. 2007. FSCB, a novel protein kinase A-phosphorylated
868 calcium-binding protein, is a CABYR-binding partner involved in late steps of fibrous
869 sheath biogenesis. J. Biol. Chem. 282:34104–34119.
870 Liang, J., and I. V. Sharakhov. 2019. Premeiotic and meiotic failures lead to hybrid male sterility
871 in the Anopheles gambiae complex. Proc. Biol. Sci. 286:20191080.
872 Liao, Y., G. K. Smyth, and W. Shi. 2014. featureCounts: an efficient general purpose program
873 for assigning sequence reads to genomic features. Bioinformatics 30:923–930.
874 Lifschytz, E., and D. L. Lindsley. 1972. The role of X-chromosome inactivation during
875 spermatogenesis. Proc. Natl. Acad. Sci. U. S. A. 69:182–186.
876 Liu, Y., J. Zhou, and K. P. White. 2014. RNA-seq differential expression studies: more sequence
877 or more replication? Bioinformatics 30:301–304.
878 Lüpold, S., G. M. Linz, J. W. Rivers, D. F. Westneat, and T. R. Birkhead. 2009. Sperm
879 competition selects beyond relative testes size in birds. Evolution 63:391–402.

880 Ma, S., A. S. Avanesov, E. Porter, B. C. Lee, M. Mariotti, N. Zemskaya, R. Guigo, A. A.
881 Moskalev, and V. N. Gladyshev. 2018. Comparative transcriptomics across 14
882 Drosophila species reveals signatures of longevity. Aging Cell 17:e12740.
883 Mack, K. L., P. Campbell, and M. W. Nachman. 2016. Gene regulation and speciation in house
884 mice. Genome Res. 26:451–461.
885 Mack, K. L., and M. W. Nachman. 2017. Gene regulation and speciation. Trends Genet. 33:68–
886 80.
887 Maekawa, M., C. Ito, Y. Toyama, F. Suzuki-Toyota, T. Kimura, T. Nakano, and K. Toshimori.
888 2004. Stage-specific expression of mouse germ cell-less-1 (mGCL-1), and multiple
889 deformations during mgcl-1 deficient spermatogenesis leading to reduced fertility. Arch.
890 Histol. Cytol. 67:335–347.
891 Manceau, M., V. S. Domingues, R. Mallarino, and H. E. Hoekstra. 2011. The developmental role
892 of Agouti in color pattern evolution. Science 331:1062–1065.
893 McCarthy, D. J., Y. Chen, and G. K. Smyth. 2012. Differential expression analysis of multifactor
894 RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40:4288–
895 4297.
896 Meiklejohn, C. D., E. L. Landeen, J. M. Cook, S. B. Kingan, and D. C. Presgraves. 2011. Sex
897 chromosome-specific regulation in the Drosophila male germline but little evidence for
898 chromosomal dosage compensation or meiotic inactivation. PLoS Biol. 9:e1001126.
899 Montgomery, S. H., and J. E. Mank. 2016. Inferring regulatory change from gene expression: the
900 confounding effects of tissue scaling. Mol. Ecol. 25:5114–5128.
901 Morgan, K., B. Harr, M. A. White, B. A. Payseur, and L. M. Turner. 2020. Disrupted gene
902 networks in subfertile hybrid house mice. Mol. Biol. Evol. 37:1547–1562.

903 Mugal, C. F., M. Wang, N. Backström, D. Wheatcroft, M. Ålund, M. Sémon, S. E. McFarlane,
904 L. Dutoit, A. Qvarnström, and H. Ellegren. 2020. Tissue-specific patterns of regulatory
905 changes underlying gene expression differences among Ficedula flycatchers and their
906 naturally occurring F1 hybrids. Genome Res. 30:1727–1739.
907 Mukaj, A., J. Piálek, V. Fotopulosova, A. P. Morgan, L. Odenthal-Hesse, E. D. Parvanov, and J.
908 Forejt. 2020. Prdm9 intersubspecific interactions in hybrid male sterility of house mouse.
909 Mol. Biol. Evol. 37:3423–3438.
910 Murat, F., N. Mbengue, S. B. Winge, T. Trefzer, E. Leushkin, M. Sepp, M. Cardoso-Moreira, J.
911 Schmidt, C. Schneider, K. Mößinger, T. Brüning, F. Lamanna, M. R. Belles, C. Conrad,
912 I. Kondova, R. Bontrop, R. Behr, P. Khaitovich, S. Pääbo, T. Marques-Bonet, F.
913 Grützner, K. Almstrup, M. H. Schierup, and H. Kaessmann. 2021. The molecular
914 evolution of spermatogenesis across mammals. bioRxiv, doi:
915 10.1101/2021.11.08.467712.
916 Namekawa, S. H., P. J. Park, L.-F. Zhang, J. E. Shima, J. R. McCarrey, M. D. Griswold, and J.
917 T. Lee. 2006. Postmeiotic sex chromatin in the male germline of mice. Curr. Biol.
918 16:660–667.
919 Newman, A. M., C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S.
920 Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, and A. A. Alizadeh.
921 2019. Determining cell type abundance and expression from bulk tissues with digital
922 cytometry. Nat. Biotechnol. 37:773–782.
923 Nguyen, T. B., K. Manova, P. Capodieci, C. Lindon, S. Bottega, X.-Y. Wang, J. Refik-Rogers, J.
924 Pines, D. J. Wolgemuth, and A. Koff. 2002. Characterization and expression of
925 mammalian cyclin b3, a prepachytene meiotic cyclin. J. Biol. Chem. 277:41960–41969.

926 Oka, A., A. Mita, Y. Takada, H. Koseki, and T. Shiroishi. 2010. Reproductive isolation in hybrid
927 mice due to spermatogenesis defects at three meiotic stages. Genetics 186:339–351.
928 Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, M. H. H. Stevens, M. J. Oksanen, and M.
929 Suggests. 2007. The vegan package. Community ecology package 10:719.
930 Pitnick, S., D. J. Hosken, and T. R. Birkhead. 2009. Sperm morphological diversity. Pp. 69–149
931 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm Biology. Academic Press,
932 London.
933 Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg,
934 M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape
935 underlying phenotypic integrity in the face of gene flow in crows. Science 344:1410–
936 1414.
937 Poelstra, J. W., N. Vijay, M. P. Höppner, and J. B. Wolf. 2015. Transcriptomics of colour
938 patterning and colouration shifts in crows. Mol. Ecol. 24:4617–4628.
939 Price-Waldman, R. M., A. J. Shultz, and K. J. Burns. 2020. Speciation rates are correlated with
940 changes in plumage color complexity in the largest family of songbirds. Evolution
941 74:1155–1169.
942 Rafati, N., J. A. Blanco-Aguiar, C. J. Rubin, S. Sayyab, S. J. Sabatino, S. Afonso, C. Feng, P. C.
943 Alves, R. Villafuerte, N. Ferrand, L. Andersson, and M. Carneiro. 2018. A genomic map
944 of clinal variation across the European rabbit hybrid zone. Mol. Ecol. 27:1457–1478.
945 Ramm, S. A., and L. Schärer. 2014. The evolutionary ecology of testicular function: size isn't
946 everything. Biol. Rev. Camb. Philos. Soc. 89:874–888.
947 Ramm, S. A., L. Schärer, J. Ehmcke, and J. Wistuba. 2014. Sperm competition and the evolution
948 of spermatogenesis. Mol. Hum. Reprod. 20:1169–1179.

949 Raymond, C. S., M. W. Murphy, M. G. O'Sullivan, V. J. Bardwell, and D. Zarkower. 2000.
950 Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis
951 differentiation. Genes Dev. 14:2587–2595.
952 Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: a Bioconductor package for
953 differential expression analysis of digital gene expression data. Bioinformatics 26:139–
954 140.
955 Rohlfs, R. V., and R. Nielsen. 2015. Phylogenetic ANOVA: the expression variance and
956 evolution model for quantitative trait evolution. Syst. Biol. 64:695–708.
957 Rolland, A. D., J.-J. Lareyre, A.-S. Goupil, J. Montfort, M.-J. Ricordel, D. Esquerré, K. Hugot,
958 R. Houlgatte, F. Chalmel, and F. Le Gac. 2009. Expression profiling of rainbow trout
959 testis development identifies evolutionary conserved genes involved in spermatogenesis.
960 BMC Genomics 10:546.
961 Sato, D. X., N. Rafati, H. Ring, S. Younis, C. Feng, J. A. Blanco-Aguiar, C.-J. Rubin, R.
962 Villafuerte, F. Hallböök, M. Carneiro, and L. Andersson. 2020. Brain transcriptomics of
963 wild and domestic rabbits suggests that changes in dopamine signaling and ciliary
964 function contributed to evolution of tameness. Genome Biol. Evol. 12:1918–1928.
965 Schärer, L., D. T. J. Littlewood, A. Waeschenbach, W. Yoshida, and D. B. Vizoso. 2011. Mating
966 behavior and the evolution of sperm design. Proc. Natl. Acad. Sci. U. S. A. 108:1490–
967 1495.
968 Schultz, N., F. K. Hamra, and D. L. Garbers. 2003. A multitude of genes expressed solely in
969 meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. Proc.
970 Natl. Acad. Sci. U. S. A. 100:12201–12206.

971 Schwahn, D. J., R. J. Wang, M. A. White, and B. A. Payseur. 2018. Genetic dissection of hybrid
972 male sterility across stages of spermatogenesis. Genetics 210:1453–1465.
973 Shen-Orr, S. S., and R. Gaujoux. 2013. Computational deconvolution: extracting cell type-
974 specific information from heterogeneous samples. Curr. Opin. Immunol. 25:571–578.
975 Shima, J. E., D. J. McLean, J. R. McCarrey, and M. D. Griswold. 2004. The murine testicular
976 transcriptome: characterizing gene expression in the testis during the progression of
977 spermatogenesis. Biol. Reprod. 71:319–330.
978 Snyder, R. L. 1967. Fertility and reproductive performance of grouped male mice. Pp. 458–472
979 *in* K. Benirschke, ed. Comparative aspects of reproductive failure. Springer, Berlin.
980 Srivatsan, S. R., M. C. Regier, E. Barkan, J. M. Franks, J. S. Packer, P. Grosjean, M. Duran, S.
981 Saxton, J. J. Ladd, M. Spielmann, C. Lois, P. D. Lampe, J. Shendure, K. R. Stevens, and
982 C. Trapnell. 2021. Embryo-scale, single-cell spatial transcriptomics. Science 373:111–
983 117.
984 Stern, D. L., and V. Orgogozo. 2008. The loci of evolution: how predictable is genetic evolution?
985 Evolution 62:2155–2177.
986 Suzuki, T. A., and M. W. Nachman. 2015. Speciation and reduced hybrid female fertility in
987 house mice. Evolution 69:2468–2481.
988 Todd, E. V., M. A. Black, and N. J. Gemmell. 2016. The power and promise of RNA-seq in
989 ecology and evolution. Mol. Ecol. 25:1224–1241.
990 Torgasheva, A. A., and P. M. Borodin. 2016. Cytological basis of sterility in male and female
991 hybrids between sibling species of grey voles Microtus arvalis and M. levis. Sci. Rep.
992 6:36564.

993 Torgerson, D. G., R. J. Kulathinal, and R. S. Singh. 2002. Mammalian sperm proteins are rapidly
994 evolving: evidence of positive selection in functionally diverse genes. Mol. Biol. Evol.
995 19:1973–1980.
996 Trapnell, C., L. Pachter, and S. L. Salzberg. 2009. TopHat: discovering splice junctions with
997 RNA-Seq. Bioinformatics 25:1105–1111.
998 Turner, L. M., E. B. Chuong, and H. E. Hoekstra. 2008. Comparative analysis of testis protein
999 evolution in rodents. Genetics 179:2075–2089.
1000 Turner, L. M., and B. Harr. 2014. Genome-wide mapping in a house mouse hybrid zone reveals
1001 hybrid sterility loci and Dobzhansky-Muller interactions. Elife 3:e02504.
1002 Turner, L. M., D. J. Schwahn, and B. Harr. 2012. Reduced male fertility is common but highly
1003 variable in form and severity in a natural house mouse hybrid zone. Evolution 66:443–
1004 458.
1005 Turner, L. M., M. A. White, D. Tautz, and B. A. Payseur. 2014. Genomic networks of hybrid
1006 sterility. PLoS Genet. 10:e1004162.
1007 Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics.
1008 Nat. Rev. Genet. 10:57–63.
1009 White-Cooper, H., K. Doggett, and R. E. Ellis. 2009. The evolution of spermatogenesis. Pp.
1010 151–183 *in* T. R. Birkhead, D. J. Hosken, and S. Pitnick, eds. Sperm Biology. Academic
1011 Press, London.
1012 Whittington, E., T. L. Karr, A. J. Mongue, S. Dorus, and J. R. Walters. 2019. Evolutionary
1013 proteomics reveals distinct patterns of complexity and divergence between Lepidopteran
1014 sperm morphs. Genome Biol. Evol. 11:1838–1846.

1015 Widmayer, S. J., M. A. Handel, and D. L. Aylor. 2020. Age and genetic background modify
1016 hybrid male sterility in house mice. Genetics 216:585–597.
1017 Wittkopp, P. J. 2007. Variable gene expression in eukaryotes: a network perspective. J. Exp.
1018 Biol. 210:1567–1575.
1019 Wohnhaas, C. T., G. G. Leparc, F. Fernandez-Albert, D. Kind, F. Gantner, C. Viollet, T.
1020 Hildebrandt, and P. Baum. 2019. DMSO cryopreservation is the method of choice to
1021 preserve cells for droplet-based single-cell RNA sequencing. Sci. Rep. 9:10699.
1022 Yapar, E., E. Saglican, H. M. Dönertaş, E. Özkurt, Z. Yan, H. Hu, S. Guo, B. Erdem, R. V.
1023 Rohlfs, P. Khaitovich, and M. Somel. 2021. Convergent evolution of primate testis
1024 transcriptomes reflects mating strategy. bioRxiv, doi: 10.1101/010553.
1025 Yoshikawa, H., D. Xu, Y. Ino, T. Yoshino, T. Hayashida, J. Wang, R. Yazawa, G. Yoshizaki,
1026 and Y. Takeuchi. 2018. Hybrid sterility in fish caused by mitotic arrest of primordial
1027 germ cells. Genetics 209:507–521.