OXFORD GENETICS

GENETICS, 2023, 223(2), iyac181
https://doi.org/10.1093/genetics/iyac181
Advance Access Publication Date: 9 December 2022
Investigation

Regulation at *Drosophila*'s *Malic Enzyme* highlights the complexity of transvection and its sensitivity to genetic background

Teresa Z. Rzezniczak,[1] Mark T. Rzezniczak,[1] Bruce H. Reed,[2] Ian Dworkin,[3] Thomas J.S. Merritt[1,*]

[1]Department of Chemistry & Biochemistry, Laurentian University, Sudbury, ON P3E 2C6, Canada
[2]Department of Biology, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[3]Department of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada

*Corresponding author: Department of Chemistry & Biochemistry, Laurentian University, Sudbury, ON P3E 2C6, Canada. Email: tmerritt@laurentian.ca

Abstract

Transvection, a type of *trans*-regulation of gene expression in which regulatory elements on one chromosome influence elements on a paired homologous chromosome, is itself a complex biological phenotype subject to modification by genetic background effects. However, relatively few studies have explored how transvection is affected by distal genetic variation, perhaps because it is strongly influenced by local regulatory elements and chromosomal architecture. With the emergence of the "hub" model of transvection and a series of studies showing variation in transvection effects, it is becoming clear that genetic background plays an important role in how transvection influences gene transcription. We explored the effects of genetic background on transvection by performing two independent genome wide association studies (GWASs) using the *Drosophila* genetic reference panel (DGRP) and a suite of *Malic enzyme* (*Men*) excision alleles. We found substantial variation in the amount of transvection in the 149 DGRP lines used, with broad-sense heritability of 0.89 and 0.84, depending on the excision allele used. The specific genetic variation identified was dependent on the excision allele used, highlighting the complex genetic interactions influencing transvection. We focussed primarily on genes identified as significant using a relaxed *P*-value cutoff in both GWASs. The most strongly associated genetic variant mapped to an intergenic single nucleotide polymorphism (SNP), located upstream of *Tiggrin* (*Tig*), a gene that codes for an extracellular matrix protein. Variants in other genes, such transcription factors (*CG7368* and *Sima*), RNA binding proteins (*CG10418, Rbp6,* and *Rig*), enzymes (*AdamTS-A, CG9743*, and *Pgant8*), proteins influencing cell cycle progression (*Dally* and *Eip63E*) and signaling proteins (*Atg-1*, *Axo, Egfr,* and *Path*) also associated with transvection in *Men*. Although not intuitively obvious how many of these genes may influence transvection, some have been previously identified as promoting or antagonizing somatic homolog pairing. These results identify several candidate genes to further explore in the understanding of transvection in *Men* and in other genes regulated by transvection. Overall, these findings highlight the complexity of the interactions involved in gene regulation, even in phenotypes, such as transvection, that were traditionally considered to be primarily influenced by local genetic variation.

Keywords: genetic background, transvection, malic enzyme

Introduction

Gene regulation is complicated and the biological impact of noncoding genetic variation can be difficult to predict, but changes in gene expression via regulatory regions of genes play a substantial role in evolution (Wray 2007). This variation includes local regulatory elements such as promoters, and other nearby *cis*-regulatory elements, as well as distal regulatory elements such as enhancers, insulators, and repressors. In *Drosophila*, the somatic pairing of homologous chromosomes further complicates gene regulation (Metz 1916; McKee 2004). The close physical proximity between homologs is involved in *trans*-regulation of the genes, in which regulatory elements on one homolog influence elements on the other, adding a layer of complexity to the already intricate system of gene regulation. This type of *trans*-regulation, termed transvection (Lewis 1954), is pairing-dependent; disruption of pairing through chromosomal inversions, for example, eliminates the regulation and can result in misregulation of expression (Tian et al. 2019).

Traditionally, transvection is studied using complementation between two specific mutations, for example, a promoter deletion and an enhancer deletion (Geyer et al. 1990). In this situation, transvection results in intragenic complementation of these alleles. This type of interaction occurs at some, but not all genes. One of these genes, *yellow*, has a single promoter and transcriptional start site, and several well-defined enhancers (Morris et al. 1998, 1999, 2004). Mutations in these elements result in the *yellow* phenotype, but flies that are heterozygous for a *yellow* allele with a deletion or disruption in the promoter and a *yellow* allele with a deletion or disruption of an enhancer have wild-type coloration, and this intragenic complementation can be disrupted by preventing chromosome pairing (Morris et al. 1999).

Received: October 12, 2022. **Accepted:** November 23, 2022

We have developed a system that allows us to study transvection using only a single mutant allele, heterozygous with wild-type chromosomes (Merritt et al. 2005; Lum and Merritt 2011; Bing et al. 2014). *Malic enzyme* (*Men*) codes for a metabolic enzyme of the same name (MEN). Flies that are homozygous for small knockout deletions in the *Men* regulatory region have no *Men* expression or MEN protein activity. In the absence of intragenic complementation, knockout heterozygotes are expected to have 50% wild-type activity and flies that are heterozygous for large deletions in *Men* meet these expectations. Strikingly, flies that are heterozygous for small knockout (null) mutations in the *Men* regulatory region have a range of enzyme activities ranging from 50% to over 100% wild-type activity (Lum and Merritt 2011). The MEN protein activity is entirely a function of the intact copy on the wild-type chromosome; the mutant chromosome is involved in regulation but is not transcribed (Lum and Merritt 2011). This trans-regulation at the *Men* locus is pairing dependent and is not physiological upregulation in response to decreased MEN activity (Bing et al. 2014), thereby making this regulation transvection, as traditionally defined. Because this system requires only a single mutant allele, it allows us to easily quantify transvection across genetic backgrounds simply by crossing a line with a small *Men* deletion to a suite of wild-type lines.

The *Men* system also allows us to accurately quantify variation in transvection. Although transvection is a gene expression effect, we can measure it in this system by quantifying enzyme activity, allowing us to resolve differences in transvection as low as 5%, far lower than can typically be measured using gene expression. This accuracy along with our abilities to easily cross *Men* deletions to a series of backgrounds and to easily manipulate environmental conditions allowed us to show that transvection is itself a complex and plastic trait sensitive to differences in local mutations, genetic backgrounds, and the environment (Bing et al. 2014). Enzyme activity and gene expression are closely correlated at the *Men* locus, so enzyme activity serves as a strong proxy, with greater resolution, for *Men* expression (Lum and Merritt 2011).

Traditionally, studies of transvection focus heavily on local genetic variation, such as differences in the location, size of deletions, and genetic variation in enhancers just upstream of the gene of interest (King et al. 2019; Piwko et al. 2019). This focus on local variation likely arose from the classical model of transvection, in which enhancers (or silencers) were thought to interact directly with the promoter *in trans*. However, a newer model of transvection suggests that transvection is mediated through the formation of transcriptional hubs, which involve an interplay of interactions between insulators, enhancers, transcription factors, and polymerases (Lim et al. 2018), rather than through a single event of direct contact between an enhancer and a promoter. This hub model of transvection opens the door for the influence of distal genetic variation on transvection and could be the underlying reason for the strong influence of genetic background on transvection in *Men* (Lum and Merritt 2011; Bing et al. 2014). For example, genetic variation in a gene that codes for a transcription factor could influence the interplay of interactions within the "hub," thereby influencing transvection.

The differential influence of distal genetic variants between genetic backgrounds with a focal allele of a gene is known as genetic background effects (GBEs). GBEs have been identified in a number of different complex traits, including enzyme activity, in which a single mutation (induced or natural) exerts different phenotypic effects depending on the genetic background in which it is placed (*e.g.* Rzezniczak and Merritt 2012; Chandler et al. 2013). Although most studies of transvection have focused largely on the influence of local genetic variation, specifically in the promoters, enhancers, silencers, and insulators adjacent to the gene of interest, other studies have shown that the amount of transvection can vary across genetic backgrounds, suggesting the influence of distal genetic variation through GBE (Lum and Merritt 2011; Bing et al. 2014). The variation across genetic backgrounds is consistent with the finding by Chandler et al. 2017 that shows GBE are greatest in phenotypes with moderate expressivity such as transvection.

In this study, we used the *Men* transvection system and a well-characterized suite of genetic backgrounds to explore GBE in transvection. We quantified the amount of transvection-mediated misregulation in *Men* across the *Drosophila* genetic reference panel (DGRP), a community resource of inbred fly lines with fully sequenced genomes, collected from a single wild population in Raleigh, NC, USA (Mackay et al. 2012). By crossing the DGRP to two *Men* promoter-disrupting alleles with no MEN activity (and a wild-type control), we generated two separate, but complimentary, data sets and performed a genome wide association study (GWAS) for each set. We identified distal genetic variation associated with transvection in *Men* that mapped to a variety of different genes, including transcription factors, RNA binding proteins, signaling proteins, enzymes, and proteins involved in cell cycle progression. This study suggests transvection is influenced by an interplay of the local and distal genetic variation, consistent with the "hub" model of transvection.

Materials and methods

Fly stocks and maintenance

149 DGRP fly stocks were used in this study and were obtained from the Bloomington *Drosophila* Stock Centre. We also used a suite of three lines bearing *P*-element-mediated excisions in the *Men* gene (*MenEx* lines) that are described elsewhere (Lum and Merritt 2011; Bing et al. 2014), but we will briefly describe them here. The *MenEx3* had a perfect excision of the *P*-element resulting in an allele with wild-type MEN activity. The two other lines had large imprecise excisions, resulting in the complete loss MEN activity: *MenEx86* had a 2239 bp deletion and *MenEx76* had a 669 bp deletion. Both deletions are located around the transcriptional start site and remove the promoter (Fig. 1) and have no MEN activity (Lum and Merritt 2011). All three excision chromosomes are isogenic outside of the excisions at the *P*-element site and were placed in the same genetic background: *w; 6326; i/TM8*, where *i* is each excision chromosome described earlier. Five male flies from each of the DGRP lines were crossed to five females from each of the three lines bearing *P*-element-mediated excisions in *Men*. The offspring collected from the cross between each of these excision lines and the DGRP lines are heterozygous at each chromosome: *w/Y; 6326/DGRP; i/DGRP*. Flies were maintained at 25°C, on a standard cornmeal–yeast–agar–corn syrup diet, with 12-hour light/dark cycles.

Sample collection and processing

Male flies from each cross were collected, under light CO_2 anesthesia, and aged three–five days. For most lines, 10 biological replicates were collected; in some cases, fewer replicates were used if the number of flies collected was limited. For each biological replicate, aged flies were collected, in groups of four, and stored at −80°C. Flies were homogenized in ice-cold grinding buffer (100 mM Tris–HCl, 0.15 mM $NADP^+$, and pH 7.4) at a "concentration" of one fly per 100 µl of the buffer. Homogenized samples

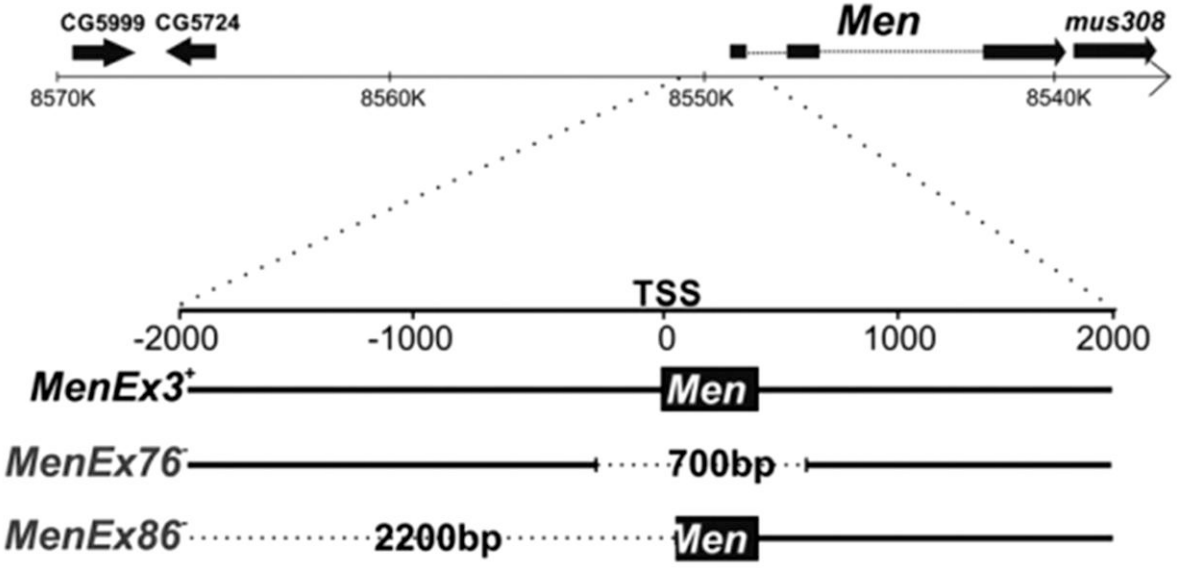

Fig. 1. *Men* gene region. The *Men* gene region showing the relative sizes of the *P*-element mediated deletions in the *MenEx*- alleles. *MenEx86* has a deletion of 2239 bp upstream of *Men*, including the promoter and some of the first exon, whereas *MenEx76* has a 669 bp deletion which removes some of the upstream region, the promoter, and all the first exon. *Men3* has a perfect excision of the *P*-element and is used as a control line. The three chromosomes bearing these three excisions are genetically identical, save for the size of the deletion (or lack thereof).

were spun at 13,000 RPM for 10 minutes at 4°C to pellet all solids, and the supernatant was collected to measure MEN activity.

MEN activity assays

MEN activity assays were performed on a Molecular Devices SpectraMax 384 Plus 96-well plate spectrophotometer, using 10 µl of fly homogenate and 100 µl of MEN-specific assay buffer (100 mM Tris–HCl, 0.34 mM $NADP^+$, 50 mM Malate, 50 mM $MnCl_2$, and pH 7.4). Absorbance was measured at 340 nm, every 9 seconds, over 3 minutes at 25°C. Each biological sample was assayed at least twice using independent aliquots. Enzyme activities are expressed as optical density (OD) units per min. MEN activity was used as a proxy for *Men* expression, as the two measures are highly correlated with $r = 0.84$ (Lum and Merritt 2011).

Estimating quantitative genetic parameters

To estimate fundamental genetic parameters (heritabilities, coefficient of genetic variation, genetic correlations), a linear mixed modeling approach was used. We fit a single model with *Men* allele (Men^3, Men^{76}, Men^{86}) as a fixed effect, and random effects were fit for the experimental block, an unstructured (3 × 3) covariance matrix for random effects of DGRP varying by *Men* allele, a diagonal [3 × 3] matrix to estimate biological sample-level variances varying by *Men* allele (along diagonal). Residual variance (reflecting residual variation for aliquots) was allowed to vary according to *Men* allele. This full model was fit using *MCMCglmm()* in the MCMCglmm library. The default uninformative improper prior was used throughout. 200,000 Iterations for the MCMC were used with a burn-in of 10,000 and thinning interval of 5. While results from MCMCglmm are shown throughout, we confirmed all results using *lmer()* in lme4 and *glmmTMB()* in glmmTMB. lme4 does not allow for fully varying residual variances by allele, nor could *Men* allele sample-level variances be fit. As such, checks of heritability estimates for each *Men* allele were confirmed by fitting allele-specific models. In all cases, estimates were extremely similar. All models were fit using R v4.0.2 on macOS Catalina 10.15.7, with the following libraries: MCMCglmm_2.32, lme4_1.1-26, glmmTMB_1.0.2.1, lmerTest_3.1-3, emmeans_1.5.5-1, pbkrtest_0.5.1 (Hadfield 2010; Bates et al. 2015; Brooks et al. 2017; R Core Team 2020; Lenth et al. 2022). Heritabilities were estimated as the ratio of the allele-specific variance among DGRP to the sum of allele-specific variation due to DRGP and allele-specific sample variances. Genetic correlations were estimated by generating vectors of correlations from the posterior distributions of the relevant variance and covariance terms. Unless otherwise specified, point estimates are the posterior means provided with 95% highest posterior density intervals.

Estimating *Men* up- or downregulation via transvection

In this study, we quantified transvection at the *Men* locus for every combination of *Men* excision allele *MenEx* $^-$ (where (–) is *MenEx86* or *MenEx76*) and DGRP line. Transvection, pairing-dependent regulation of gene expression, at the *Men* locus results in MEN activity that is a nonlinear combination of the parent alleles when one of the alleles is a promoter-deficient knockout deletion. There is essentially no transvection when both *Men* alleles are wild type (Bing et al. 2014), but each wild-type line has its own characteristic MEN activity reflecting line-specific differences in gene expression (Baath and Merritt 2019; Effect A in Fig. 2). To account for these line-specific differences, we used the MEN enzyme activity in *MenEx3/DGRP#* flies, where *DGRP#* represents one of the 150 DGRP lines used in this study, as a no-transvection baseline for each cross. Because the *MenEx3*, *MenEx86*, and *MenEx76* alleles are identical except for the deletions, MEN enzyme activity for each *MenEx86/DGRP#* and *MenEx76/DGRP#* cross differs from the *MenEx3/DGRP#* baseline only by the amount of transvection in each combination (Fig. 2). As such, the ratio of MEN activity in *MenEx*$^-$*/DGRP#* to *MenEx3/DGRP#* indicates the presence and magnitude of transvection. The use of ratio variables is problematic for statistical inference in general (and were not used for inference in this study). This use is doubly problematic as the effects of *Men76* and *Men86* alleles were both scaled by the wild-type *Men3* allele effects. While the ratio variables were not used for the

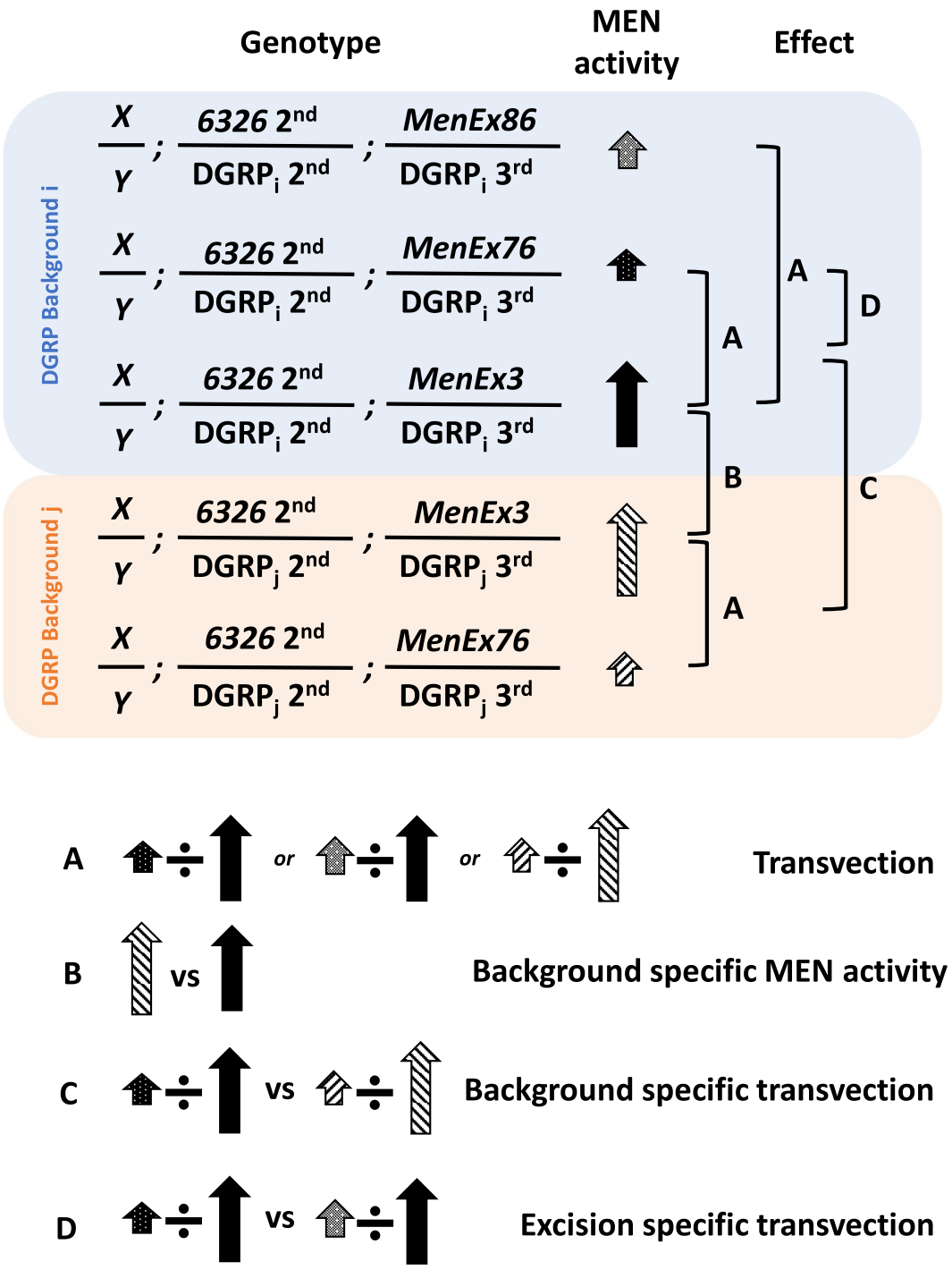

Fig. 2. *Men* model of transvection summarized for two genetic backgrounds. We crossed lines containing two *Men* excision alleles, *Men76* and *Men86* (collectively *MenEx*$^{-}$ alleles) to 149 DGRP lines. Two of those DGRP lines are shown in this figure: DGRP background i and DGRP background *j*. The blue box represents the three crosses we performed for a single DGRP background: one cross to each of the *MenEx-* alleles (represented by smaller arrows to and one cross to a line containing *MenEx3*). We repeated these crosses for each DGRP background, although only two of these crosses are represented in this figure for DGRP background *j*. The lines containing *MenEx3* are genetically identical to the lines carrying the *MenEx-* alleles, except that *MenEx3* has an intact *Men* regulatory region, and are used as a baseline to account for differences in MEN activity unrelated to transvection (Effect B). Transvection (Effect A) is calculated as a ratio of the activity in the lines containing *MenEx*$^{-}$alleles to those containing to the line containing *Men3*. We found variation in the amount of transvection across backgrounds (Effect C) when using the same excision allele. In many backgrounds, we also found that the excision itself interacted with the background to influence the amount of transvection (Effect D). In other words, a high transvection background in the *Men76* data set might be a low transvection background in the *Men86* data set.

estimation of any genetic parameters, for ease of interpretation they are plotted and presented. These ratios were calculated from the allele-specific conditional means for each DGRP line. The ratios were directly computed for each iteration of the MCMC of the primary model (which directly incorporated genetic correlations in the fit), and the posterior means and 95% highest posterior density intervals were generated for each DGRP ratio. This ratio was used as a measure of transvection, expressed as a

percentage, of MEN activity found for each *MenEx*$^{-}$/DGRP# to its matching *MenEx3*/DGRP# cross (B in Fig. 2). A ratio of 50% represents a lack of transvection (i.e. pairing a promoter-deficient *Men* allele with a wild-type *Men* allele from the DGRP should result in 50% activity of the activity found when pairing the wild-type *Men* allele from *Men3* with the same DGRP line in the absence of pairing-dependent genetic interactions). MEN activity over 50% indicates the presence of pairing-dependent interactions that positively impact gene expression, whereas MEN activity under 50% indicates the presence of pairing-dependent interactions that negatively impact gene expression.

Linkage disequilibrium pruning and principal component analysis

We downloaded the DGRP genome data set (freeze 2), containing SNP, insertion, and deletion information for all 149 DGRP lines tested from the DGRP website (Mackay et al. 2012). We performed LD pruning using PLINK v 1.90 (Purcell et al. 2007) to avoid identifying genetic variations associated with our phenotypes due to linkage disequilibrium (LD), using a simple pairwise threshold (indep-pairwise 50 5 0.5). The top five principal components (PCs) were identified using SmartPCA software in Eigensoft v. 6.14 (no outlier exclusion; Patterson et al. 2006).

Genome-wide association

Genome-wide association was performed for transvection using the "—linear" flag in PLINK v 1.90 (Purcell et al. 2007); 2,778,429 SNPs were tested, following removal of SNPs with a minor allele frequency (MAF) < 2%. In addition to the focal SNP in the model, the following covariates were included as predictors. To account for the influence of residual population structure, and cryptic relatedness, the first five PCs (EigVal1-5) from the PCA analysis on LD pruned genome-wide polymorphism data were included (as earlier). To account for the influence of known chromosomal inversions, we generated a PCA for dummy variable coding of inversion status, and the first 5 PCs (inversionPC1-5) were also used as covariates. Dummy coding of *Wolbachia* infection status was used as a covariate. To account for natural variation in MEN activity not related to transvection, we included dummy codings for two nonsynonymous *Men* SNPs known to influence MEN activity, as well as continuous MEN activity level from each DGRP line crossed to *w*; *6326*; *MenEx3* as covariates. As such, for the h^{th} (focal) SNP, the linear model takes the form of

$$y_i = \beta_0 + \beta_{h,1}x_{h,i1} + \sum_{j=2}^{15}\beta_{h,j}x_{ij} + \epsilon_{h,i}$$

where y_i is the conditional mean for MEN expression from either *MenEx86/DGRP*$_i$ or *MenEx76/DGRP*$_i$ derived from the mixed model described earlier. β_0 is the model intercept, and $\beta_{h,1}$ is the estimated effect for the model for SNP$_h$. The remaining model coefficients ($\beta_{2,h} - \beta_{15,h}$) are associated with the effects of the 14 common predictors described earlier. These estimates are expected to be similar across models (SNP to SNP), except for when there is a strong correlation between the focal SNP and one or more of these predictors. $\epsilon_{i,h}$ represents unmodeled (residual) variation. We used a Bonferroni-corrected significance threshold of $P < 1.78 \times 10^{-8}$ and a more relaxed threshold of $P < 10^{-4}$. The F1 males used in this study inherited their X chromosome from the *Men* excision lines, which allowed us to use quantile–quantile (QQ) plots to test whether the SNPs from the X chromosome followed an expected null distribution.

To evaluate whether SNPs or genes identified between *Men76* and *Men86* GWAS data sets were more common than expected by chance alone, we conducted simple simulations assuming the independence of sites and/or genes. We performed three separate analyses, each consisting of 1,000 simulations, and observed how many variants would be common between the random variants or genes. In the first set of simulations, 237 and 588 SNPS were randomly selected from the 2,778,429 evaluated SNPs in each simulation. Next, 237 and 588 SNPs were randomly selected from 1,050,279 SNPs after LD pruning. Last, 163 and 313 genes were randomly chosen from 15,094 annotated autosomal genes. The number of variants or genes randomly selected was chosen to correspond to the number of significant variants or genes in the *Men76* and *Men86* data sets.

Results

We quantified variation in transvection at the *Men* locus in flies heterozygous for a *Men* knockout allele and a *Men* functional allele using a pair of *Men* knockout chromosomes and a suite of DGRP wild-type chromosomes (DGRPi). The knockout alleles, *MenEx76* and *MenEx86* (*MenEx-*), are different excisions of the promoter and a portion of the upstream regulatory region of the *Men* locus (Fig. 1) that result in complete loss of MEN activity in homozygotes (Lum and Merritt 2011). Flies bearing these alleles have had their X and second chromosomes replaced and are genetically identical, except in the size of the deletion in *Men* (Lum and Merritt 2011). These flies were each individually crossed to 149 DGRP lines (298 crosses total), generating two distinct, but complimentary, data sets of MEN activity across the DGRP lines: hereafter the *Men76* data set and the *Men86* data set. Different DGRP fly lines have different wild-type amounts of MEN activity (background-specific MEN activity, Effect B in Fig. 2), so we also crossed flies with a functional *Men* allele (*MenEx3*) to the same set of DGRP lines, establishing a baseline of MEN activity in the presence of two functional *Men* alleles for each DGRP line. Flies with the *MenEx3* allele are genetically identical to flies bearing the *MenEx*$^{-}$ alleles, except that *MenEx3* has an intact *Men* regulatory region. The relative MEN activity for each DGRP line was quantified as the ratio of MEN activity in the *MenEx-*/DGRPi heterozygotes to that in *MenEx3*/DGRPi heterozygotes (Effect A in Fig. 2). The *MenEx*$^{-}$ alleles do not generate a functional transcript of *Men* (Lum and Merritt 2011); *Men* expression in *MenEx-*/DGRPi heterozygotes is only from the DGRP chromosome. Therefore variation in the amount of MEN activity across genetic backgrounds is caused by modification of the functional allele's expression. Based on previous work, we expected to find variation in the amount of transvection across backgrounds within each data set (background-specific transvection, Effect D in Fig. 2). By generating two data sets, we were also able to examine how the two different *MenEx*$^{-}$ alleles influenced transvection within the same genetic background (excision-specific transvection, Effect C in Fig. 2).

Genetic background strongly influences transvection at *Men* through complex interactions with the *MenEx*$^{-}$ alleles

Genetic background had a very strong effect on the relative amount of MEN activity. We fitted a linear mixed model to estimate broad-sense heritability (H^2), coefficient of genetic variation (CV_G), and the genetic correlations among MenEx alleles (Table 1). H^2 for each allele was high, indicating a strong influence of genetic variation on differences in MEN activity across the DGRP. We

Table 1. Coefficient of genetic variation (CV_G) and broad-sense heritabilities (H^2) determined from between-line variances for each *MenEx* allele. Posterior means with 95% credible intervals (HPD).

	MenEx3	*MenEx86*	*MenEx76*
CV_G (HPD)	0.178 (0.141–0.230)	0.248 (0.206–0.299)	0.240 (0.198–0.290)
H^2 (HPD)	0.816 (0.752–0.890)	0.887 (0.853–0.924)	0.840 (0.795–0.892)

observed a modest increase in H^2 and CV_G for the *MenEx76* and *MenEx86* alleles compared to wild-type *MenEx3* (Table 1). This increase was confirmed using a parametric bootstrap based on *lmer* model fits where the model allowing for *MenEx* allele-specific genetic variances was always the better fitting model among 1,000 simulations and with an asymptotic likelihood ratio test (LR = 1875, $df = 5$, $P < 2e^{-16}$). Figure 3 shows genetic correlations (among strains of the DGRP) across *MenEx* alleles were moderate with the wild-type *MenEx3* allele, with $r_{g\text{-}3,76} = 0.68$ (0.54–0.82 CI), $r_{g\text{-}3,86} = 0.65$ (0.49–0.80 CI), and the genetic correlation between mutant alleles being high $r_{g\text{-}76,86} = 0.87$ (0.81–0.92 CI). We fit the model allowing for allele-specific sample variances. We observed modest changes across alleles, with *MenEx76* allele being slightly elevated: $CV_{Men3} = 0.083$ (0.078–0.088), $CV_{Men76} = 0.10$ (0.096–0.11), and $CV_{Men86} = 0.088$ (0.081–0.095).

The relative MEN activity ranged from 39 to 99% of wild-type activity for the *Men76* dataset, and 35 to 96% for the *Men86* dataset (Fig. 4). Only the flies with approximately 50% relative MEN activity completely lack transvection, i.e. in only those flies is there no pairing-dependent up- or downregulation of *Men* expression. All other flies had transvection-based up- or downregulation of *Men* expression and relative MEN not equal to 50%. To clarify the relationship between the relative MEN activity and transvection, we created a transvection value term, such that backgrounds without any transvection, i.e. those with a relative MEN activity of 50%, will have a transvection value of zero (Fig. 5). Flies with pairing-dependent upregulation, i.e. relative MEN activity of more than 50%, have positive transvection, with transvection values ranging from greater than zero to one. A transvection value of one indicates a complete rescue of wild-type MEN activity through pairing-dependent upregulation of the single functional *Men* allele. On the other hand, flies with pairing-dependent downregulation, i.e. relative MEN activity of less than 50%, have negative transvection with transvection values ranging from less than zero to negative one. A value of negative one would be a complete lack of MEN activity, a condition that is theoretically possible but was not observed in this experiment.

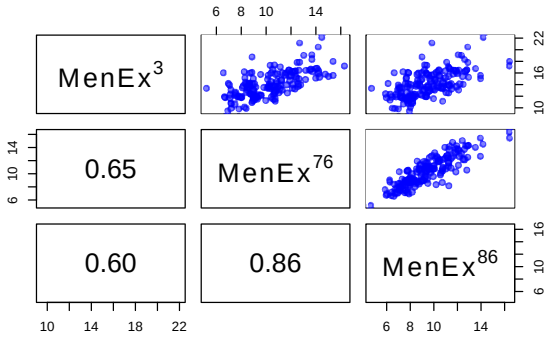

Fig. 3. Genetic correlations for MEN activity between MenEx alleles across DGRP backgrounds. Genetics correlations between *Men3* and each of the two *MenEx*$^-$ alleles (*Men76* and *Men86*) were moderate with $r_{g\text{-}3\text{–}76} = 0.61$ and $r_{g\text{-}3\text{–}86} = 0.56$. The correlation for MEN activity between *MenEx*$^-$ alleles was stronger with $r_{g\text{-}76\text{-}86} = 0.86$. Points represent condition means for each DGRP strain.

Most lines had positive transvection effects on *Men*. In the *Men76* data set, 94% of the backgrounds had positive transvection, with three of the backgrounds having a transvection value of approximately +1 (we used a transvection value range of +0.9 to +1.1 to account for experimental error), indicating a complete rescue of wild-type MEN activity. Similarly, in the *Men86* data set, 84% of the backgrounds had positive transvection with two backgrounds having a transvection value of approximately +1. DGRP555 was the single background that had a transvection value of +1 in both datasets.

Some genetic backgrounds had no evidence of transvection in *Men*. Twenty-three of the 149 genetic backgrounds had no transvection at the *Men* locus in at least one of the two data sets, with seven of these backgrounds having no transvection in both datasets. To account for experimental error, we defined "no transvection" as backgrounds that had a MEN activity ratio of 0 ± 0.1 Previous studies have shown that transvection-lacking backgrounds, like the ones identified in this study, carry chromosomal inversions that hinder transvection through their disruption of chromosomal pairing, although the relationship between inversions and loss of transvection can be complicated (Bing et al. 2014). Several DGRP lines have been previously annotated for the presence of inversions (Huang et al. 2014), and inversions of chromosome 3R (where *Men* is located) were identified in some of the DGRP backgrounds with "no transvection." Four of these backgrounds (DGRP031, DGRP136, DGRP309, and DGRP559) had an annotated inversion, *In(3R)K*, whose breakpoints encompass the *Men* gene. DGRP309, another background that showed no transvection in the *Men86* dataset but had a MEN transvection value of +0.22 in the *Men76* data set, also carries this inversion. Transvection at the *Men* locus has been shown to be disrupted by inversions of chromosome 3R (Bing et al. 2014), so the lack of transvection in flies with these backgrounds could be a result of less chromosomal pairing due to the inversion. There was one other background of the 23 "no transvection" backgrounds with an annotated chromosomal inversion of chromosome 3R: DGRP776 which carries *In(3R)P*. This inversion has breakpoints that do not encompass *Men*, but the presence of these large inversions may, nevertheless, influence chromosomal pairing, thereby influencing transvection in *Men* (Bing et al. 2014).

The remaining 18 backgrounds lacking transvection lack did not have annotated inversions on Chromosome 3R, but the absence of transvection could be the result of inversions that were not tested for in the characterization of the DGRP lines, so we selected one of the other genetic backgrounds that showed no transvection and performed a polytene chromosomal squash to visually check for large-scale inversions. This genetic background, DGRP399, did not show evidence of any chromosomal inversions (Supplementary Fig. 1). Chromosomal squashes, however, can fail to resolve small inversions, and the lack of transvection in DGRP399, and the other backgrounds lacking annotated inversions in 3R, could be the result of small, and probably local, inversions.

Very few DGRP lines had negative transvection effects. While there is substantial variation in the amount of positive transvection across backgrounds, two and only two backgrounds, DGRP757 and

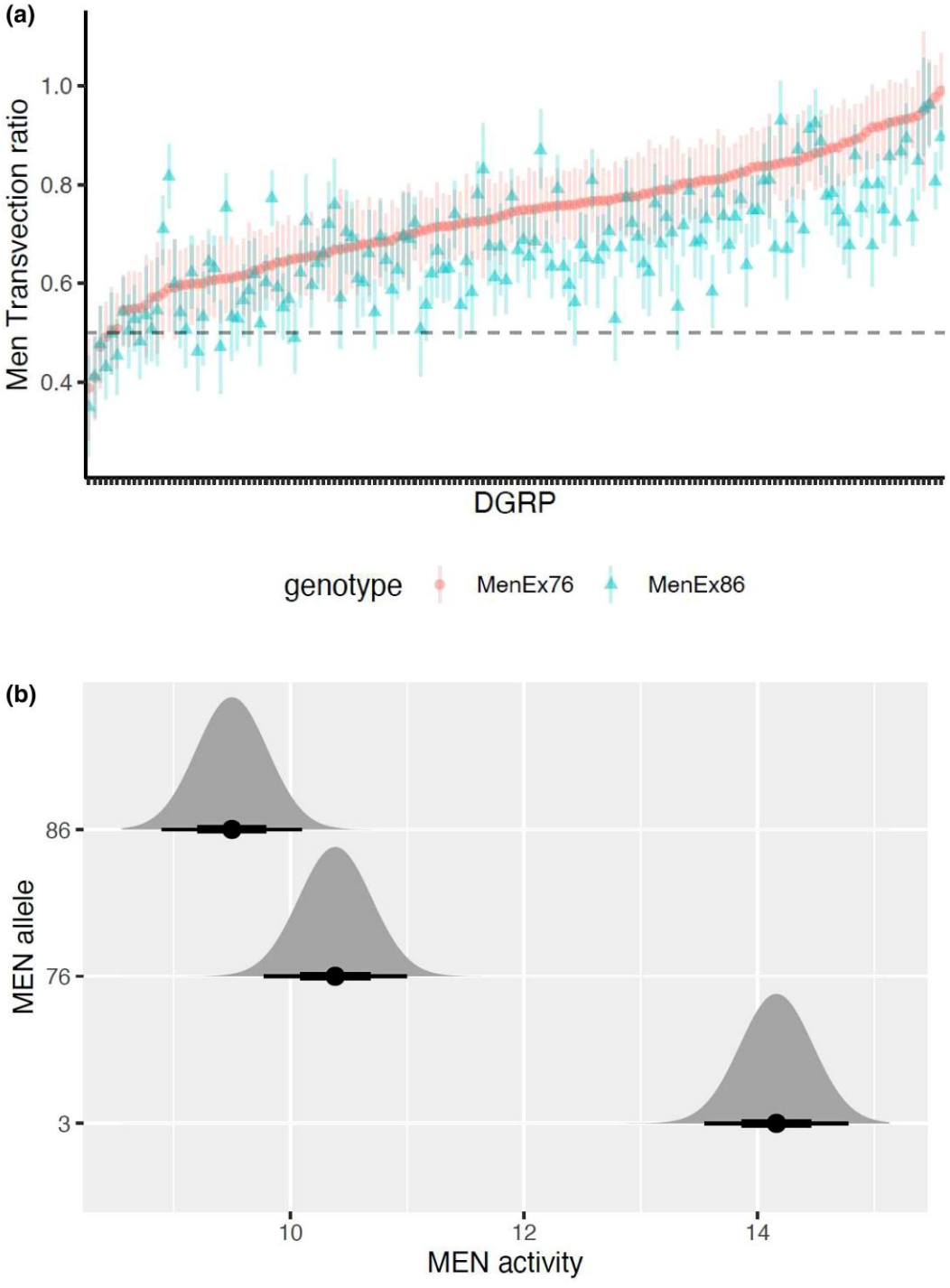

Fig. 4. Relative MEN activity across DGRP lines is correlated between data sets. a) The MEN activity for each DGRP line in the *Men76* and *Men86* data sets is represented as a ratio of MEN activity relative to that of the *Men3*/DGRP heterozygotes (wild-type controls with no transvection). In the *Men76* data set (shown in pink) MEN activity varied from 39 to 99% of wildtype, whereas in the *Men86* dataset, MEN activity varied from 35 to 96%. The dotted line represents MEN activity in cases where no transvection occurs. Although some backgrounds showed no evidence of transvection, most backgrounds had a MEN activity ratio over 0.5, indicating some level of transvection, with some backgrounds showing a full recovery of wild-type activity. Although MEN activity is correlated between the data sets, there are backgrounds that show stark differences depending on which excision allele was used. Error bars represent the 95% credible intervals (as shown in Materials and methods). b) Estimated marginal means (from the mixed model) of MEN activity across 149 backgrounds for each *Men* allele. The full extent of black lines represents 95% confidence intervals. Plotted *t*-distribution on estimates is provided to visualize uncertainty in estimates. Overall, MEN activity is similar between the *Men76* and *Men86* datasets; however, there are excision-specific differences between the backgrounds.

DGRP884, showed *Men* downregulation via transvection (negative transvection) in both data sets, with transvection values between −0.30 and −0.17. One other background, DGRP357, showed negative transvection only with *MenEx86*, but not *MenEx76*, where the background showed no transvection. It is unclear what about these three genetic backgrounds is responsible for the negative

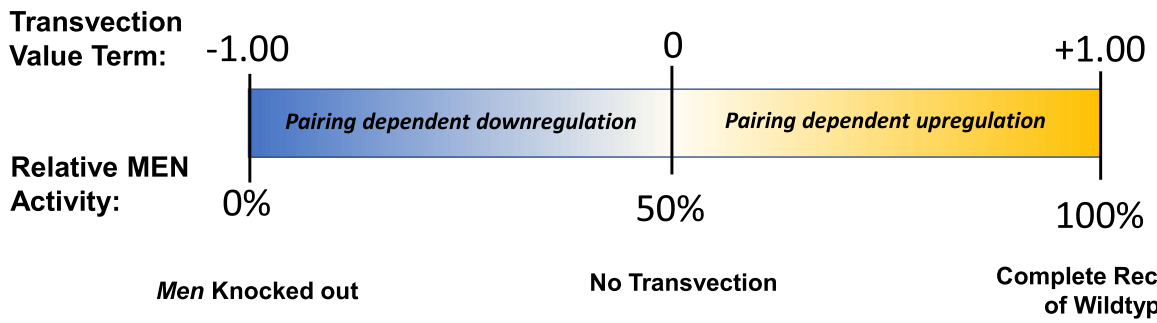

Fig. 5. Transvection value term. Transvection in the *Men* system is calculated as a ratio to account for natural variation in the amount of MEN activity between DGRP lines. This ratio is referred to as relative MEN activity. In order to make it more clear how this ratio translates to the amount of transvection at *Men*, we created a transvection value term, where a relative MEN activity of 50%, representing backgrounds without transvection, would have a transvection value term of zero. Any backgrounds showing pairing-dependent upregulation of MEN activity would have a positive transvection value term, with a value of +1.00 representing complete recovery of wild-type MEN activity. Flies with pairing-dependent downregulation of MEN activity would have a transvection value term of less than zero (negative numbers).

transvection: the region surrounding *Men* does not contain any identified or putative polycomb response elements (PREs), which have been shown to silence genes via transvection in other studies (Kassis 2002; Tian et al. 2019). However, there are also cases in which pairing-dependent silencing is associated with sequences of DNA that are not PREs (Kassis 2002), although these sequences are difficult to identify. Generally, the identification of silencers has lagged behind that of enhancers (Johnson et al. 2015) which may explain why no putative silencers have been identified in the *Men* region, leaving open the idea that the negative transvection in these lines could be a function of silencers, albeit unidentified ones. It is noteworthy that so few lines show negative transvection and that there appears to be less variability in negative transvection, although with a small sample size. Across the three backgrounds that have negative transvection effects, the amount of transvection differed very little between the $MenEx^{-}$ alleles. Even for DGRP357, where *MenEx86* showed negative transvection but *MenEx76* did not, the difference in MEN activity was only 6% between the two alleles within this background. This finding suggests that when a silencing element is present, it likely has an overwhelming effect on transvection levels, either eliminating transvection completely or resulting in negative transvection.

Overall, both $MenEx^{-}$ alleles resulted in positive transvection when paired with most DGRP backgrounds. Further levels of transvection were similar in most $MenEx^{-}$ allele/background combinations with a coefficient of correlation (r) of 0.86 between data sets. However, some DGRP backgrounds have very different levels of transvection with either excision allele. For example, DGRP721 showed no transvection value of +0.06 (no transvection) when paired with *MenEx76*, but a transvection value of +0.55 (strong transvection) when paired with *MenEx86*, indicating strong upregulation. These kinds of differences reinforce our previous results from this system of strong interactions between the genetic background and the size or location of the deletion in the *Men* regulatory region (Lum and Merritt 2011; Bing et al. 2014) and highlight the intricacy of the interactions which govern transvection and gene regulation in general.

Use of two data sets allows for identification of candidate genes associating with MEN activity despite potential population stratification in one of the datasets

A large amount of variation in transvection across the DGRP backgrounds makes this regulatory phenomenon an excellent phenotype for GWAS, and we were able to identify genetic variation which correlates with transvection; however, most hits were identified using a relaxed significance cutoff (Figs. 6 and 7). We used transvection at the *Men* locus as the quantitative trait to perform single marker regression in both the *Men76* and *Men86* data sets. A total of 2.8 million SNPs were tested, following the removal of SNPs with a MAF of less than 2%. We included several covariates in the analysis to avoid confounding the GWAS: MEN activity in the DGRP line when crossed to wild-type *Men3* chromosome, the genotype of the two nonsynonymous polymorphisms known to influence MEN activity, the top five PCs from the PCA to avoid possible population stratification and cryptic relatedness, the top five PCs from a PCA of the annotated chromosomal inversions in the DGRP, which could influence chromosomal pairing (and thus transvection) and *Wolbachia* infection status which can cause cryptic relatedness and bias GWAS (Ivanov et al. 2015).

QQ plots were generated using the *P*-values from the GWAS for both data sets (Fig. 8) to test each for population stratification which can confound the GWAS if population structure correlates with the phenotype. Separate QQ plots were created for the autosomes (Fig. 8, a and b) and the X chromosome (Fig. 8, c and d) data because the flies used in the GWAS were F1 males that inherited their X chromosome from the line bearing the *Men* excision allele rather than from the DGRP. Because this X chromosome did not come from the GWAS set, we expected to find fewer associations between transvection and X-linked SNPs in the GWAS (He et al. 2014). QQ plots for the *Men76* data set show systematic inflation of *P*-values from expected indicative of left-skewed data. This type of shift can be indicative of population stratification. However, the QQ plot for the *Men86* data is normally distributed, with little separation between observed and expected *P*-values. Since we utilized the same set of DGRP lines in both GWAS, it is unlikely that the left-skewed data in the *Men76* data set is caused by population stratification as we do not see this same pattern of unusual differentiation in the *Men86* data set, another strength of using two data sets.

SNP upstream of *Fic/Tig* is the only significant variant associating with MEN activity using Bonferroni-corrected *P*-value

The GWAS analysis for both data sets found very few significant SNPs associated with transvection in *Men* using a Bonferroni-corrected *P*-value cutoff of 0.05 ($P < 1.78 \times 10^{-8}$). The

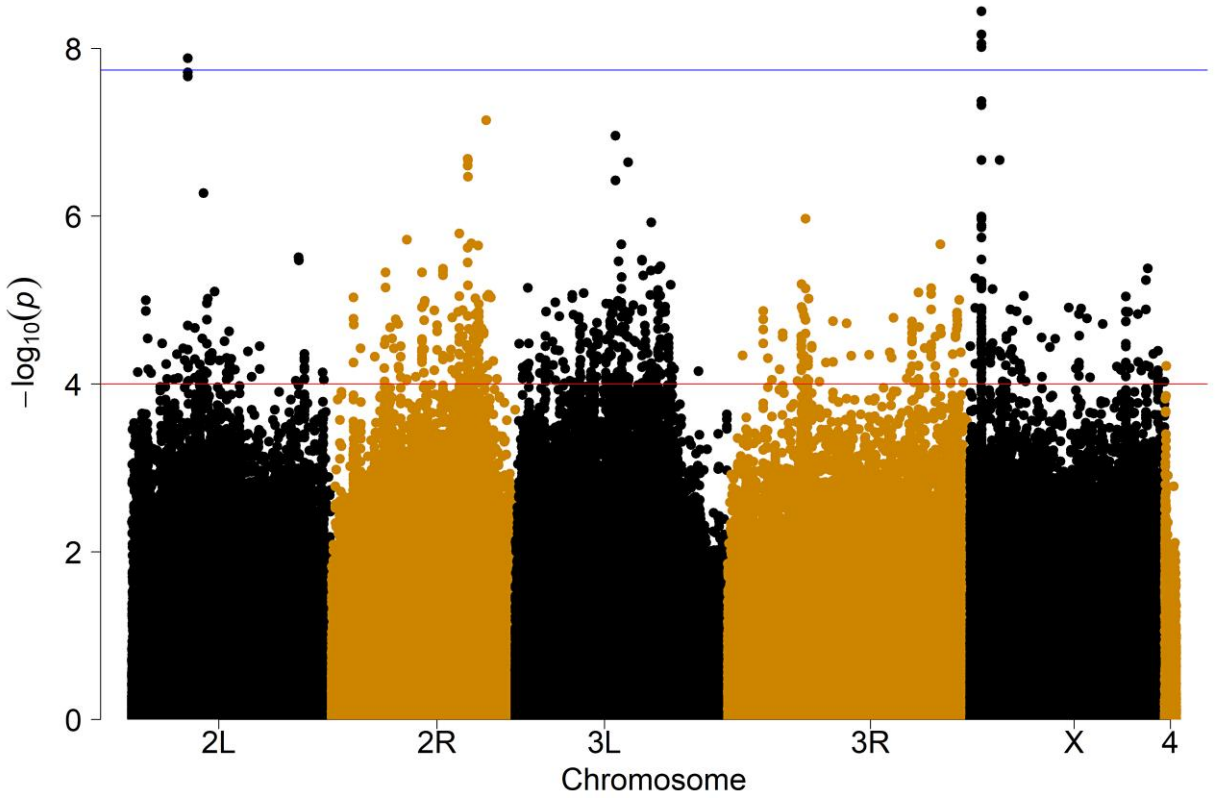

Fig. 6. Genome-wide analysis for the *Men76* data set identifies a single strong candidate locus associated with MEN activity. This locus, located on Chromosome 2L, was significant using a Bonferroni-corrected cutoff of $P < 0.05$ (shown in red). The locus maps to an SNP located in the intergenic region between the genes *Fic* and *Tig*. The blue horizontal line indicates a more relaxed *P*-value cutoff of 10^{-4}. The association peak on chromosome X indicates the presence of false positives in the data set as the X chromosome in the flies (which were Male) was not inherited from the DGRP data set and was used as a negative control.

Men86 data set did not identify any variants that reached this threshold (Fig. 7). However, using a relaxed *P*-value threshold of 10^{-4} to identify further genes of interest, which may influence transvection in *Men*. At this threshold, the *Men86* data set had 237 significant autosomal variants in the *Men86* data set, with peaks of interest on the Manhattan plot (Fig. 7; Supplementary Table 1). Twenty-three of the significant variants mapped to a large intergenic space on Chromosome 3L, downstream of the gene *ventral veins lacking* (*vvl*), which codes for a transcription factor involved in several processes, including cell fate determination. In addition, the top variant in the *Men86* data set mapped to *Beaten path IIIA* (*Beat-IIIa*), is predicted to be involved in cell-cell adhesion. Eight additional variants were associated with *Beat-IIIa*, all of which were intronic.

In the *Men76* data set, there was a single SNP that broke the Bonferroni-corrected *P*-value cutoff (Fig. 6). This SNP is found in an intergenic region between *Fic* and *Tig* at a location 320 bp upstream of *Fic* and 706 bp upstream of *Tig*. The peak at Chromosome 2 L on the Manhattan plot (Fig. 6) shows there are several other top hits in this same intergenic region. This same variant upstream of *Fic/Tig* was identified as one of the top hits in the *Men86* data set. *Tig* was also identified in previous work conducted by Joyce et al. (2012), which used a fluorescent in situ hybridization approach to conduct an RNA screen that identified pairing-promoting and anti-pairing genes. Reduction of *Tig* expression by RNAi was found to reduce pairing in one of the two cell lines tested in the paper (Joyce et al. 2012). *Tig* codes for an extracellular matrix glycoprotein and integrin ligand (Fogerty et al. 1994).

Using a relaxed *P*-value cutoff for the *Men76* data set, 588 significant autosomal variants were associated with MEN activity in the *Men76* data set (Fig. 6; Supplementary Table 2). Because the QQ plots suggested population stratification in the *Men76* data set, we focussed mostly on genes that were identified as significant in both data sets. Overall, there were 26 significant autosomal variants that were common to both data sets. In addition, we assigned variants to specific genes if they were located within the gene or within 1000 bp of a gene in either the 5′ or 3′ direction. We then identified genes that had significant variants within both data sets even if the specific variants varied between the data sets. Using this approach, we identified a total of 33 genes or gene regions (in the case that a variant was between genes or multiple genes overlapped; Table 2).

We performed simulations to determine how many variants would be found in common between two random data sets of similar size, to lend credibility to the common genes identified in the *MenEx*$^{-}$ data sets. We randomly selected 237 and 588 variants (independent of each other), of the 2,778, 429 variants evaluated. Of 1,000 simulations, we found a maximum of three shared sites (95% quantile of 0) between the two randomly selected data sets, compared to 26 observed shared variants in the *MenEx*$^{-}$ data sets. The 1,000 simulations performed after LD pruning, also showed a maximum of 3 shared sites (95% quantile of 1 shared variant), with 11% of the simulations having any shared sites.

Perhaps, most pertinent and biologically relevant, at the gene level 33, genes were shared across the two data sets. In contrast, when 163 and 313 genes were randomly selected from the

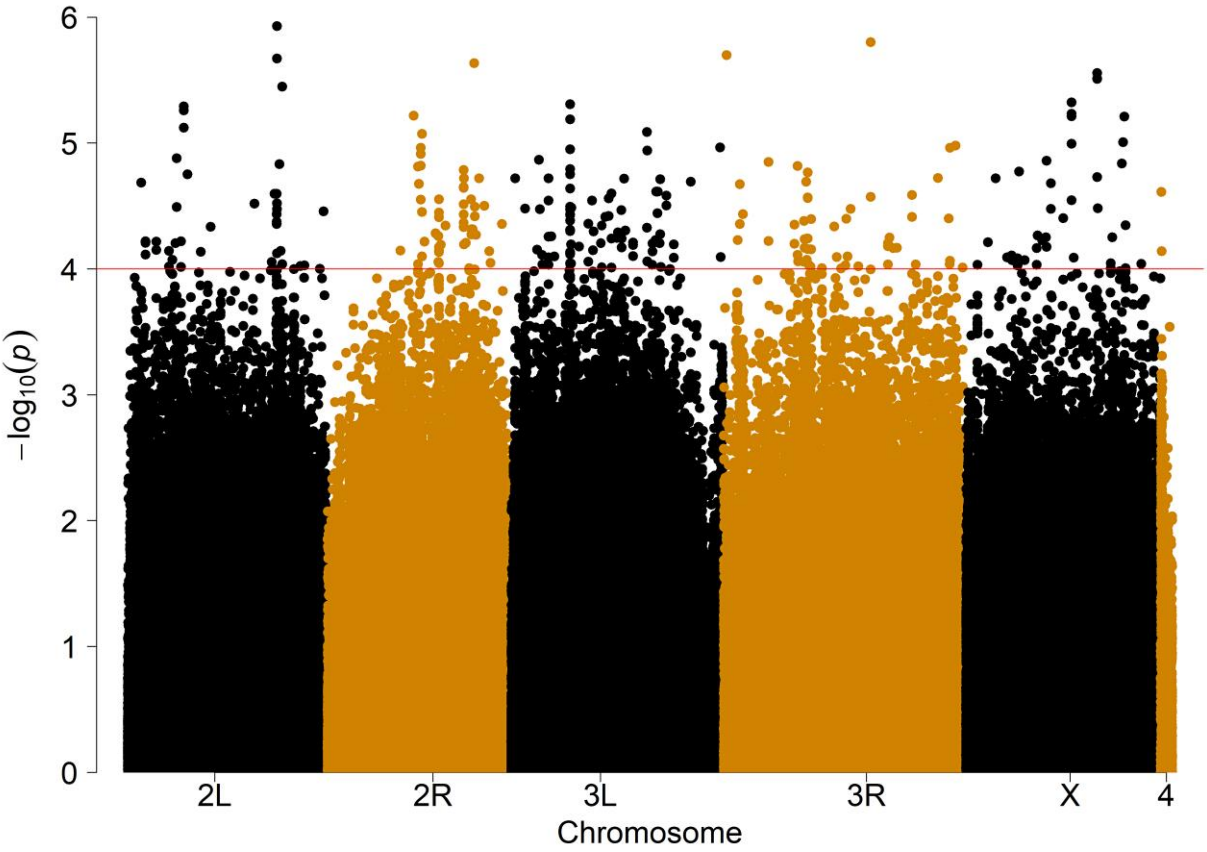

Fig. 7. Genome-wide analysis for MEN activity in the *Men86* data set identifies several candidate genes using a relaxed *p*-value cutoff. There is a distinct peak at Chromosome 3L which maps to an intergenic region downstream of the transcription factor *vvl*. In addition, many of the top hits on Chromosome 2L map to an intergenic region between genes *Fic* and *Tig* which was a locus identified as the top hit in the *Men76* dataset as well. In addition, several variants associated with a gene on *Beat-IIIa* (Chromosome 2L), including the top hit in the data set.

annotated autosomal genes (15, 094), we observed a maximum of 12 shared genes (95% quantile of 6 shared genes) out of the 1,000 simulations. Although the simulations do not account for LD structure or correlations among response variables (independent of a relationship between response variables and SNPs), they do lend credibility to the variants and genes that were found to be common in the *MenEx*$^{-}$ datasets as future targets of study.

GWAS identifies variants in genes coding for metabolic proteins

We expected to find evidence of genetic variants influencing transvection at *Men* in one of two general categories: (1) variants that influence transvection in *Men* only, which could include variants in the genes that code for transcription factors that directly influence *Men* expression, and (2) variants that influence chromosomal pairing in general and could influence transvection universally. Of the 33 genes identified as significant in both data sets, it appears that there are genes that fit into both categories.

Many of the genes associated with transvection in *Men* in both data sets (Table 2) were enzymes involved in varying cellular processes. Given MEN's role in metabolism, it may be that the products of these genes may associate with transvection in *Men* specifically. Some of the identified common genes that code for enzymes include *AdamsTS-A*, *Axo*, *Pgm2b*, *CG17292*, *CG4467*, and *CG9743/CG15531* (Table 2). It is likely that these enzymes associate with transvection through indirect methods. For example, feedback from a metabolic process in which an enzyme is involved could influence cellular levels of a transcription factor that acts, directly or indirectly, on the transcription of *Men* in *trans*. Some of the genes identified in the GWAS code for metabolic enzymes involved in processes downstream of MEN activity. The gene products of *CG9743* and *CG15531* are involved in the fatty acid synthesis, which consumes nicotinamide adenine dinucleotide phosphate (NADPH), one of the products of the reaction catalyzed by MEN.

In addition, variations associated with genes involved in hypoxia (*Sima*) or metal response (*Mtf1*) were also identified in both GWAS data sets (Table 2). It is also possible that these genes associate with transvection directly at *Men*, considering that MEN is a metabolic enzyme, rather than influencing chromosomal pairing in general. MEN plays a role in NADPH production for reactive oxygen species detoxification (Xiao and Loscalzo 2020), with a study in mice adipocytes showing that hypoxia influences the amount of NADPH produced by MEN (Liu et al. 2016). Whatever the interactions, they are not simply upregulating MEN activity to maintain homeostasis as previous work has shown that larger scale deletions (Lum and Merritt, 2011) or inversions (Bing et al. 2014) that eliminate transvection do not show *Men* upregulation.

We also found a single instance of variation directly upstream of *Men* that is associated with variation in transvection in the *Men76* data set ($P > 9.8 \times 10^{-5}$). This SNP, located 397 bp upstream of *Men*, is likely a variation in an unannotated transcription factor binding site (TFBS).

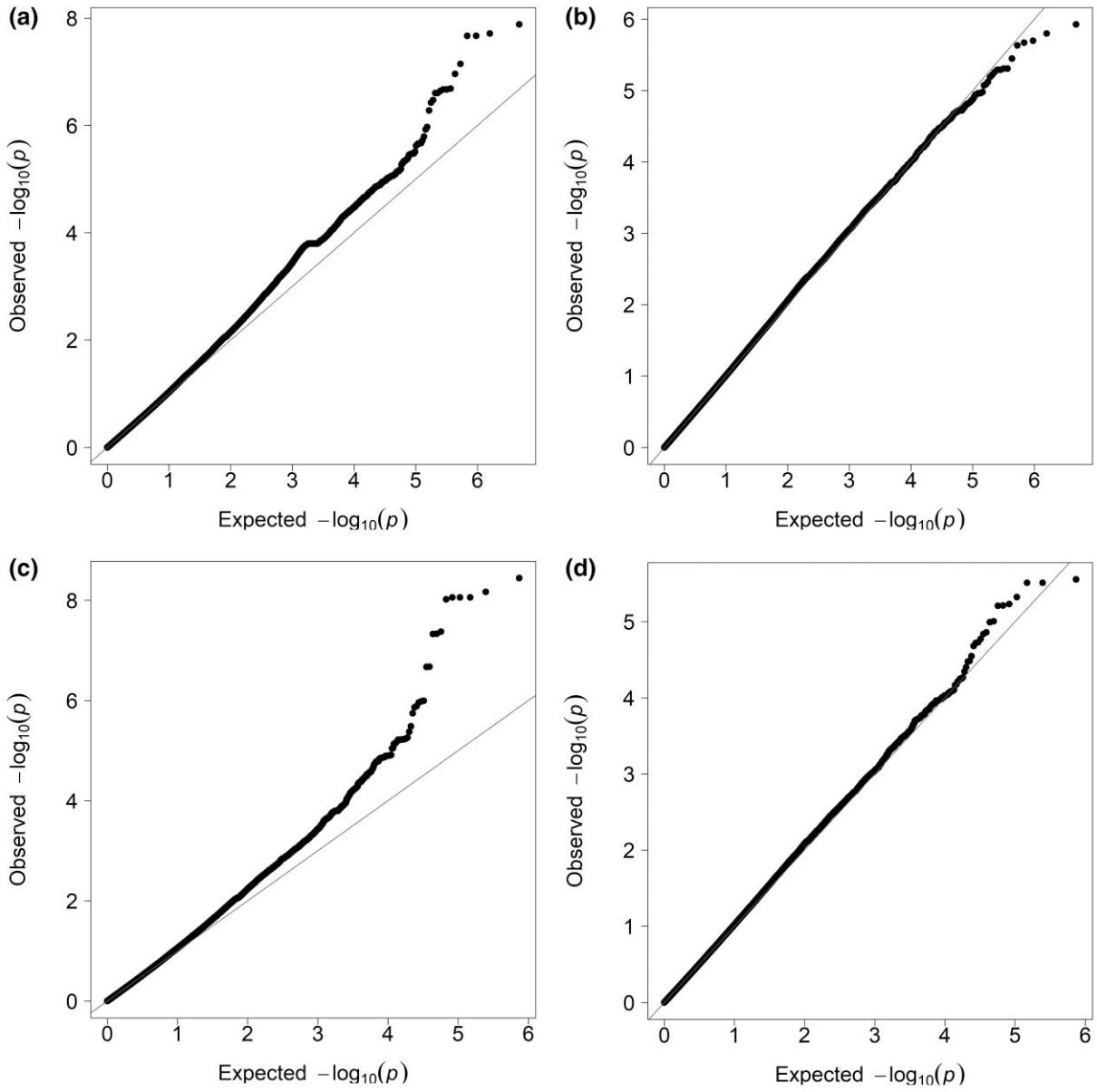

Fig. 8. QQ Plots for *Men76* and *Men86* data sets. The QQ plots for variants on the autosomes in the (A) *Men76* and (B) *Men86* data sets with the *x*-axis displaying the expected -log *P*-values for a null distribution, and the *y*-axis displaying the experimental -log *P*-values from the GWAS. The QQ plot for the variants in autosomes in *Men76* shows that there is some evidence of population stratification in the *Men76* data set; however, this data set also shows some unusually differentiated markers that are good candidates for association with MEN activity. The QQ plots for the X chromosomes in both the (C) *Men76* and (D) *Men86* data sets are shown as these were used as negative controls since the flies used were homozygous for DGRP chromosomes at chromosomes 2 and 3; however, they inherited their X chromosome from the fly lines bearing the *MenEx*$^{-}$ alleles. The QQ plot for X chromosome variants in *Men76* is consistent with the presence of population stratification in the data set.

GWAS suggests genes involved in transcription, signaling, and the cell cycle play role in mediating transvection

On the other hand, the GWAS also identified genes that likely associate with transvection, in general, perhaps by influencing chromosomal pairing. Several variants of interest in the GWAS were linked to genes involved in transcription including transcription factors (*CG7368* and *Sima*), RNA binding proteins (*CG10418, Rbp6*) and an RNA polymerase (*Tplus3b*). Additionally, some of the genes identified in both datasets play a role in cell cycle progression (*Dally* and *Eip63E*) and signaling (*Atg-1*, *Axo, Egfr,* and *Path*). These genes could influence chromosomal pairing, affecting transvection universally.

Comparing the *MenEx*$^{-}$ data sets with the candidate pairing-promoting and anti-pairing genes generated by Joyce et al. (2012) for homologous chromosomes in *D. melanogaster* cells, we found six genes that overlapped. *Tig*, mentioned earlier, was identified in both data sets and by Joyce *et al.*; however, it is not immediately evident how an extracellular matrix protein may influence chromosomal pairing and/or the amount of transvection in *Men*. The five genes were only found in one of the two data sets. These genes include *Pins*, *Raw*, *Velo*, *Pbl*, and *Pvr*. *Partner of inscuteable* (*Pins*) is the hit with the largest effect size in the *Men76* data. *Pins* is a protein that is required for *insucuteable* to asymmetrically localize during asymmetric cell division and miotic

Table 2. Genes identified with genetic variation associating significantly with MEN activity in both data sets using a relaxed *P*-value cutoff.

Gene(s)	Gene role
AdamsTS-A	Enables metalloendopeptidsase activity
Atg1	Protein kinase—autophagy regulation downstream of Tor
Axo	Serine-type endopeptidase activity
Pgm2b	Phosphopentomutase activity
CG17292	Enables lipase activity
Tplus3b	Enables RNA polymerase II C-terminal domain phosphoserine bind activity
CG4467	Metalloaminopeptidase activity; zinc binding
CG7368	Negative regulation of transcription
CG9313	Sensory perception of sound
CG9743/CG15531	Stearoyl-CoA 9-desaturase—unsaturated fatty acid synthesis Enables iron ion binding/stearoyl-CoA 9-desaturase
Con	Cell adhesion protein
Cpr51A/CG30197	Structural component of larval cuticle/enables peptidase inhibitor activity
CG10418/CR45444	RNA binding activity/long noncoding RNA gene
Dally	Coreceptor for growth factors and morphogens such as products of *dpp*, *wg*, etc.
Egfr	Transmembrane tyrosine kinase receptor
Eip63E	Cyclin-dependent kinase
Klar	Links microtubule motors with cellular components
Lrt	Transmembrane protein expressed in tendon cells
Mir-252/CG14374/ CR17025	Involved in mRNA-mediated gene silencing/unknown/antisense long noncoding RNA gene
Msp300	Positioning of cellular components in muscle cells
MTF-1	Zinc protein finger involved in the heavy metal response
Nsf2	Recycles SNARE complex proteins
Oatp30B	Transmembrane transporter of ouabain
Obp56g	Odorant binding—involved in reproduction
Obp58c	Odorant binding—involved in sensory perception
Path	Amino acid transporter. Regulates growth (TOR signaling)
Pgant8/CG7579	Catalyzes the first step of O-linked oligosaccharide biosynthesis/inactive glycosyltransferase
Pkc53E	Kinase. Involved in hh-signaling pathway. Receptor for phorbol esters
Rbp6	RNA binding protein. Role in fungal infection/aggressive behavior
Rig/CG13436	Required for motor activity. Nuclear receptor interacting protein/involved in regulation of cilium beat frequency
Sick	Cytoskeletal protein
Sima	Transcriptional regulator of hypoxia response. Binds to hypoxia response element
Tig/Fic	Extracellular matrix protein. Accumulates at embryonic and larval muscle attachment site/enzyme that catalyzes AMP residue addition to proteins

spindle orientation (Yu et al. 2000). Similarly, *Pebble* (*Pbl*) was identified in the *Men76* data set and is a gene required for cytokinesis (Lehner 1992). Both these genes play an important role in mitotic division, and variants in these genes may play a role in the pairing of homologous chromosomes, which is essential for transvection to occur.

Discussion

Over the past decade, the importance of GBE has come to the forefront of biology, in part because of the advent of fully curated genomic resources, such as the DGRP. While our understanding of the causal genetic mechanisms behind GBE remains limited, these effects appear to involve a complex interplay between a mutation, multiple genes, and the environment (Mullis et al. 2018). Still, despite its importance on the manifestation of phenotype, a potential influence of GBE is underexplored for some phenotypes, where the focus has been on local allelic factors. Pairing-dependent gene misregulation, or transvection, is one such trait, in which studying the influence of local factors (e.g. size and location of lesions) has taken precedence over distal factors (genetic background). Nevertheless, studies of transvection at *Men* have shown that transvection is in fact a plastic trait that is influenced by genetic background (Lum and Merritt 2011; Bing et al. 2014). Here, we quantified transvection at *Men* across the DGRP to demonstrate that GBE in transvection are associated with an interplay of distal genetic variation in genes whose products have a wide variety of roles in the cell. The genetic variation we identified, through GWAS, appears to influence transvection both at *Men* directly, and peripherally (i.e. by influencing general pairing-dependent mechanisms, which would in turn affect transvection at *Men*). These findings stress the importance of considering GBE in studies of transvection at other loci.

Transvection is a plastic trait that varies across genetic backgrounds

We found striking similarities and differences in transvection across the DGRP backgrounds for each *Men* excision data set. In both data sets, there was substantial variation in the amount of transvection across genetic backgrounds. Several backgrounds showed no transvection, but most of the backgrounds showed upregulation via transvection, and only two backgrounds showed downregulation via transvection. Between the two data sets, there was a strong correlation in the amount of transvection in specific genetic backgrounds suggesting that transvection at *Men* is mediated by general pairing mechanisms (i.e. distal genetic variation). Despite the strong correlation, there were still backgrounds that behaved very differently depending on the *Men* excision. This finding suggests that transvection also depends on the interaction between the genetic background and local variation (the size or location of the excision in *Men*). Interestingly, in most studies of transvection, including ours, upregulation via transvection occurs much more commonly than downregulation (Tian et al. 2019). Gene expression involves a complex interplay between multiple enhancers, silencers, and insulators. It may be that although silencers exert their influence in *trans*, their overall effect is swamped out due to the concurrent influence of enhancers *in trans*. The study of silencers has generally lagged behind the study of enhancers (Johnson et al. 2015), and

as we begin to understand these elements more thoroughly, we may also find solutions to this puzzle of the predominance of up-, and not down-, misregulation through transvection. Interestingly, one other study that explicitly examined how wild-type genetic background influenced allelic interactions, including alleles that participate in transvection, found less evidence for background-specific effects (Chandler et al. 2017). This study used complementation-based approach, as opposed to the approach used here, and only included two distinct wild-type backgrounds for the transvection analysis. As a result, the study by Chandler et al. (2017) may not have captured either the broader context of transvection or a broad distribution of naturally derived genotypes.

Influence of distal genetic variation on transvection can be explained by the hub model

In the hub model of transvection, TFBS, both in *cis* and in *trans*, are part of a complex that also associates with transcription factors, co-activators, promoters, as well as DNA polymerases and can interact with the DNA sequence on both chromosomes to influence the expression of a gene (Lim et al. 2018). Through the hub, the genetic variation on the DGRP chromosome could influence the way in which enhancers (and silencers) on the promoter-deficient chromosome interact with the promoter. Whatever the mechanism for the association, the misregulation of MEN activity is not simply reflective of physiological upregulation as feedback to low MEN activity: previous studies using a promoter-deficient allele in *Men*, which also had a large excision of the entire *Men* regulatory region, showed a complete lack of pairing-dependent up- or downregulation, suggesting that enhancers acting in *trans* are required for the misregulation (Lum and Merritt 2011; Bing et al. 2014).

We identified distal genetic variation which associated with transvection at *Men*, including variation in transcription factors, RNA binding proteins, and other elements (Table 2). The influence of this distal genetic variation fits well with the hub model of transvection, in which the amount of transvection is affected by complex interactions between all proteins (transcription factors and RNA binding proteins) and DNA sequences involved in transcription.

The fact that associations between transvection in *Men* and distal genetic variation could be mediated through interactions between multiple proteins and/or genes may explain the wide variety of functions in the genes identified through the GWAS (Table 2). For example, it is unclear how variation in a gene such as *Tig,* which codes for an extracellular matrix protein, could influence transvection in *Men* (Table 2). However, the influence of *Tig* on transvection could be mediated through an entire network of interactions. Even though the mechanism of action for some of the genes associated with transvection in *Men* is unclear, they should not be ruled out as false positives before confirmatory studies.

In addition to transcription factors, the GWAS also identified variations in RNA binding proteins that are associated with transvection. RNA binding proteins are not commonly associated with influencing transvection but could play a role, especially in the context of transvection hubs, whereby interactions between regulatory elements, transcription factors, and polymerases could be further influenced by proteins that bind pre-mRNA being transcribed.

Transvection is associated with small-effect variants that may influence transvection indirectly

The results from the GWAS of both data sets, with only a single strongly associating variant in one of the datasets, suggest that transvection itself is a polygenic trait which is consistent with the amount of transvection identified across DGRP backgrounds (Fig. 4). This finding is similar to a GWAS conducted for transvection at *spineless* found only a single, strongly associating variant, while the remaining variants showed weaker associations despite a phenotype that varied widely across the DGRP (Anderson et al. 2017). Together, these studies suggest that transvection may be a complex trait influenced by multiple other genes, some of which may be specific to the regulation of transvection while others may be specific to the regulation of the gene itself.

Identifying variants that have small effects on a phenotype, like transvection, is difficult. Crouch and Bodmer (2020) assert that small-effect variants "are likely to exert their influences on phenotypes through very indirect mechanisms, far downstream from their proximal functions, giving rise to a large number of potential candidate functions to investigate." Consistent with this assertion, our study found associations between transvection, and genes involved in processes that have not been previously associated with transvection. For example, in the *Men86* data set, gene ontology analysis found a 12.75-fold enrichment for genes involved in segment polarity determination whereas cell adhesion and regulation of neurogenesis were enriched (7.39-fold) in the *Men76* data set (data not shown). It is not immediately apparent how variation in genes involved in these processes might affect transvection. However, because transvection is a complex process, seemingly influenced by many downstream mechanisms, the association of these variants with transvection cannot be ruled out.

Our model of transvection, like many others, does not allow for confirmatory RNAi experiments, as it requires the use of *Men* excision lines, which can make differentiating variants affecting transvection through indirect mechanisms from false positives particularly challenging. However, by generating two independent data sets (*Men76* and *Men86*), we could identify common variants, thereby increasing the likelihood that these variants do influence transvection, despite the lack of confirmatory tests.

Overall, we identified a genetic variation that associates with transvection either through chromosomal pairing in general or specifically in *Men*. The distal genetic variation identified as associated with transvection is consistent with the hub model of transvection and shows how extensive the GBE on transvection can be. The GBEs that influence the amount of transvection are complex and seem to be mediated by intricate, interacting networks. This works stresses the importance of considering multiple genetic backgrounds in order to lend greater biological relevance to studies of transvection.

Data availability

The SNP/Indel information for the DGRP (Freeze 2.0) is publicly available at http://dgrp2.gnets.ncsu.edu/data.html. All other data and code are available at https://github.com/DworkinLab/TR2022_TransvectionDataScripts & https://doi.org/10.6084/m9.figshare.21605598.v1 (static DOI).

Supplemental material is available at *GENETICS* online.

Funding

This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant and a Canada Research Chair to T.J.S.M.

Conflicts of interest

The authors declare no conflict of interest.

Literature cited

R Core Team. R: A Language and Environment for Statistical Computing. Vienna (Austria): R Foundation for Statistical Computing; 2020.

Anderson C, Reiss I, Zhou C, Cho A, Siddiqi H, Mormann B, Avelis CM, Deford P, Bergland A, Roberts E, *et al.* Natural variation in stochastic photoreceptor specification and color preference in Drosophila, (S. G. Sprecher, Ed.). eLife. 2017;6:e29593. doi:10.7554/eLife.29593.

Baath S, Merritt TJS. Biological effects of single-nucleotide polymorphisms in the *Drosophila melanogaster* malic enzyme locus. Biochem Genet. 2019;58(1):129–156. doi:10.1007/s10528-019-09932-2.

Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48. doi:10.18637/jss.v067.i01.

Bing X, Rzezniczak TZ, Bateman JR, Merritt TJS. Transvection-based gene regulation in Drosophila is a complex and plastic trait. G3 (Bethesda). 2014;4(11):2175–2187. doi:10.1534/g3.114.012484.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Skaug HJ, Mächler M, Bolker BM. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 2017;9(2):378–400. doi:10.32614/RJ-2017-066.

Chandler CH, Chari S, Dworkin I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet. 2013;29(6):358–366. doi:10.1016/j.tig.2013.01.009.

Chandler CH, Chari S, Kowalski A, Choi L, Tack D, DeNieu M, Pitchers W, Sonnenschein A, Marvin L, Hummel K, *et al.* How well do you know your mutation? Complex effects of genetic background on expressivity, complementation, and ordering of allelic effects. PLoS Genet. 2017;13(11):e1007075. doi:10.1371/journal.pgen.1007075.

Crouch DJM, Bodmer WF. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. Proc. Natl. Acad. Sci. 2020;117(32):18924–18933. doi:10.1073/pnas.2005634117.

Fogerty FJ, Fessler LI, Bunch TA, Yaron Y, Parker CG, Nelson RE, Brower DL, Gullberg D, Fessler JH. Tiggrin, a novel Drosophila extracellular matrix protein that functions as a ligand for Drosophila alpha PS2 beta PS integrins. Dev. Camb. Engl. 1994;120(7):1747–1758. doi:10.1242/dev.120.7.1747.

Geyer PK, Green MM, Corces VG. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. EMBO J. 1990;9(7):2247–2256. doi:10.1002/j.1460-2075.1990.tb07395.x.

Hadfield JD. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J Stat Softw. 2010;33(2):1–22. doi:10.18637/jss.v033.i02.

He BZ, Ludwig MZ, Dickerson DA, Barse L, Arun B, Vilhjálmsson BJ, Jiang P, Park S-Y, Tamarina NA, Selleck SB, *et al.* Effect of genetic variation in a drosophila model of diabetes-associated misfolded human proinsulin. Genetics. 2014;196(2):557–567. doi:10.1534/genetics.113.157800.

Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF *et al.* Natural variation in genome architecture among 205 *Drosophila melanogaster* genetic reference panel lines. Genome Res. 2014;24(7):1193–1208. doi:10.1101/gr.171546.113.

Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TFC, Partridge L, Thornton JM. Longevity GWAS using the Drosophila genetic reference panel. J Gerontol A Biol Sci Med Sci. 2015;70-(12):1470–1478. doi:10.1093/gerona/glv047.

Johnson WC, Ordway AJ, Watada M, Pruitt JN, Williams TM, Rebeiz M. Genetic changes to a transcriptional silencer element confers phenotypic diversity within and between Drosophila species. PLOS Genet. 2015;11(6):e1005279. doi:10.1371/journal.pgen.1005279.

Joyce EF, Williams BR, Xie T, Wu C. Identification of genes that promote or antagonize somatic homolog pairing using a high-throughput FISH–based screen. PLOS Genet. 2012;8(5):e1002667. doi:10.1371/journal.pgen.1002667.

Kassis JA. Pairing-sensitive silencing, polycomb group response elements, and transposon homing in Drosophila. Adv Genet. 2002;46:421–438. doi:10.1016/s0065-2660(02)46015-4.

King TD, Johnson JE, Bateman JR. Position effects influence transvection in *Drosophila melanogaster*. Genetics. 2019;213(4):1289–1299. doi:10.1534/genetics.119.302583.

Lehner CF. The pebble gene is required for cytokinesis in drosophila. J. Cell Sci. 1992;103(4):1021–1030. doi:10.1242/jcs.103.4.1021.

Lenth RV, Buerkner P, Giné-Vázquez I, Herve M, Jung M, Love J, Miguez F, Riebl H, Singmann H. emmeans: estimated marginal means, aka least-squares means 2022 CRAN Date/Publication 2022-12-06 19:00:02 UTC.

Lewis EB. The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster*. Am Nat. 1954;88(841):225–239. doi:10.1086/281833.

Lim B, Heist T, Levine M, Fukaya T. Visualization of transvection in living Drosophila embryos. Mol. Cell. 2018;70(2):287–296.e6. doi:10.1016/j.molcel.2018.02.029.

Liu L, Shah S, Fan J, Park JO, Wellen KE, Rabinowitz JD. Malic enzyme tracers reveal hypoxia-induced switch in adipocyte NADPH pathway usage. Nat Chem Biol. 2016;12(5):345–352. doi:10.1038/nchembio.2047.

Lum TE, Merritt TJS. Nonclassical regulation of transcription: interchromosomal interactions at the malic enzyme locus of *Drosophila melanogaster*. Genetics. 2011;189(3):837–849. doi:10.1534/genetics.111.133231.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM *et al.* The *Drosophila melanogaster* genetic reference panel. Nature. 2012;482(7384):173–178. doi:10.1038/nature10811.

McKee BD. Homologous pairing and chromosome dynamics in meiosis and mitosis. Biochim. Biophys. Acta. 2004;1677(1–3):165–180. doi:10.1016/j.bbaexp.2003.11.017.

Merritt TJS, Duvernell D, Eanes WF. Natural and synthetic alleles provide complementary insights into the nature of selection acting on the men polymorphism of *Drosophila melanogaster*. Genetics. 2005;171(4):1707–1718. doi:10.1534/genetics.105.048249.

Metz CW. Chromosome studies on the Diptera. II. The paired association of chromosomes in the Diptera, and its significance. J Exp Zool. 1916;21(2):213–279. doi:10.1002/jez.1400210204.

Morris JR, Chen J, Filandrinos ST, Dunn RC, Fisk R, Geyer PK, Wu C. An analysis of transvection at the yellow locus of *Drosophila melanogaster*. Genetics. 1999;151(2):633–651. doi:10.1093/genetics/151.2.633.

Morris JR, Chen J, Geyer PK, -ting Wu C. Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis. Proc Natl Acad Sci U S A. 1998;95(18):10740–10745. doi:10.1073/pnas.95.18.10740.

Morris JR, Petrov DA, Lee AM, Wu C-T. Enhancer choice in cis and in trans in *Drosophila melanogaster*: role of the promoter. Genetics. 2004;167(4):1739–1747. doi:10.1534/genetics.104.026955.

Mullis MN, Matsui T, Schell R, Foree R, Ehrenreich IM. The complex underpinnings of genetic background effects. Nat Commun. 2018;9(1):1–10. doi:10.1038/s41467-018-06023-5.

Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLOS Genet. 2006;2(12):e190. doi:10.1371/journal.pgen.0020190.

Piwko P, Vitsaki I, Livadaras I, Delidakis C. The role of insulators in transgene transvection in Drosophila. Genetics. 2019;212(2):489–508. doi:10.1534/genetics.119.302165.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–575. doi:10.1086/519795.

Rzezniczak TZ, Merritt TJS. Interactions of NADP-reducing enzymes across varying environmental conditions: a model of biological complexity. G3 (Bethesda). 2012;2(12):1613–1623. doi:10.1534/g3.112.003715.

Tian K, Henderson RE, Parker R, Brown A, Johnson JE, Bateman JR. Two modes of transvection at the eyes absent gene of drosophila demonstrate plasticity in transcriptional regulatory interactions in cis and in trans. PLOS Genet. 2019;15(5):e1008152. doi:10.1371/journal.pgen.1008152.

Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8(3):206–216. doi:10.1038/nrg2063.

Xiao W, Loscalzo J. Metabolic responses to reductive stress. Antioxid Redox Signal. 2020;32(18):1330–1347. doi:10.1089/ars.2019.7803.

Yu F, Morin X, Cai Y, Yang X, Chia W. Analysis of partner of inscuteable, a novel player of drosophila asymmetric divisions, reveals two distinct steps in inscuteable apical localization. Cell. 2000;100(4):399–409. doi:10.1016/S0092-8674(00)80676-5.

Communicating editor: A. Long