Natural variation in copper tolerance in *Drosophila melanogaster* is shaped by transcriptional and physiological changes in the gut

Llewellyn Green[1], Santiago Radio[1], Gabriel E. Rech[1], Judit Salces-Ortiz[1], and Josefa González[1]

[1]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

Corresponding authors:
Josefa González
josefa.gonzalez@ibe.upf-csic.es

Llewellyn Green
Llewellyn.green@ ibe.upf-csic.es

ABSTRACT

Increases in industrialisation and anthropogenic activity have resulted in an increase in pollutants released into the environment. Of these pollutants, heavy metals such as copper are particularly concerning due to their bio-accumulative nature. Due to its highly heterogeneous distribution and its dual nature as both an essential micronutrient and toxic element, the genetic basis of copper tolerance is likely shaped by a complex interplay of physiological and environmental factors. *Drosophila melanogaster*, a long-standing sentinel of environmental toxins, is uniquely suited for the study of copper tolerance in arthropods and other more diverse species. In this study, we utilized the natural variation present in multiple populations of *D. melanogaster* collected across Europe to screen for variation in copper tolerance, which we found to be highly variable both within and between locations. While these collection locations covered a wide range of atmospheric and soil pollution levels, the degree of urbanization at the collection sites, rather than any other combination of environmental factors, was linked to copper tolerance. Moreover, differential expression analysis revealed that metabolism, reproduction, and protease induction contribute to copper response in tolerant and sensitive lines to different degrees. Additionally, the greatest transcriptomic and physiological responses to copper toxicity were seen in the midgut; where preservation of gut acidity is strongly linked to greater tolerance. Overall, our study provides a unique perspective on the genetic and environmental factors that shape copper tolerance in natural *D. melanogaster* populations and identifies new genes and physiological traits involved in this complex phenotype.

Introduction

Rapid industrialization and urbanization has had adverse impacts on biodiversity across ecosystems. Of the contaminants released into the environment due to increases in human activity, heavy metals are of a particular concern due to their ability to bio-accumulate in soils. Specifically with regards to copper, anthropogenic sources are thought to have a greater influence on topsoil concentrations than either lithological or geographic factors (Panagos et al. 2018). Human sources of copper are characterized by many point sources of contamination, which has resulted in a highly heterogeneous environmental distribution (Romic and Romic 2003), even across relatively short geographic distances (Orgiazzi *et al.* 2017). Due to its highly heterogeneous distribution, and its dual nature as both an essential micronutrient and toxic element, the genetic basis to copper tolerance has the potential to be shaped by a complex interplay of environmental and physiological factors.

As a commensal species, *Drosophila melanogaster* has a well-documented history as a sentinel of environmental toxins and can be readily sampled from a wide range of geographic locations, making it a prime choice species for the study of copper stress response (Wilson 2005). *D. melanogaster* has also served as an important tool in the characterisation of copper homeostasis and copper-related diseases (Navarro and Schneuwly 2017, Calap-Quintana et al 2017). As copper acts as an essential micronutrient at low doses, but can produce free radicals and damage DNA in excess, the mediation of copper often involves a complex system of regulators, chaperones, and transporters that are commonly found conserved across a wide range of species. Genetic manipulation of *D. melanogaster* has been used to successfully characterise the roles of the common metal-responsive transcription factor-1 (MTF-1) (Zhang et al 2001), *marvolio*, and the *Crt1* family of transporters which mediate copper uptake (Turski and Thiele 2007; Southon *et al.* 2008), the *ATP7* transporter, which regulates copper efflux. (Norgate *et al.* 2006), and the cysteine-rich metallothioneins, which serve as metal chaperones. (Egli *et al*, 2006; Yepiskoposyan *et al*, 2006). Like in many other small diptera, copper accumulates in the mid-gut as the fly ages, which alters gut physiology (Cioffi 1984). Once copper crosses the gut endothelium, it is sequestered by the metallothioneins in the morphologically distinct copper cells and deposited in insoluble granules in the lysozymes (McNulty *et al.* 2001). Despite the name, copper cells are considered 'cuprophobic' and are inhibited by excess copper (Dubreuil 2004). They are also responsible for stomach acid secretion, a function that is lost with age or gut damage, leading to an increase in pH (Li *et al.* 2016).

However, we cannot trust that these same alleles or processes that control copper homeostasis in genetically modified backgrounds are going to be shared in natural populations. To date, there have been several studies exploring the nature of copper tolerance in natural strains of *D. melanogaster,* both with regards to individual genes (Maroni *et al.* 1987*;* Catálan *et al* 2016) and broader developmental and learning and memory processes (Polkki and Rantala 2021, Zamberlan et al 2020).

Recently, Everman *et al* (2021) took benefit of a combination of high throughput genomic and transcriptomic approaches to uncover a number of new gene copper candidates, using recombinant inbred lines. They found that copper resistance is genetically complex and impacted by variation in copper avoidance behavior. In addition to identifying natural variants involved in response to copper, the pairing of genomic data with transcriptomic data also provides a greater opportunity to identify factors that regulate copper induced changes in expression, beyond the well-known MTF-1 factor. Prior expression analysis on metal exposure suggest that there are a number of co-regulated gene clusters linked to broader stress and metabolism related pathways in response to heavy metal exposure, independent of MTF-1 (Merritt and Bewick 2017, Roelofs et al 2009, Everman 2021). However, the factors responsible for these co-ordinated changes in expression have not yet been identified.

To date, genome-wide studies investigating the genetic basis of tolerance to copper and other heavy metals in *D. melanogaster* have focused on SNP variants (Everman et al 2021, Zhou et al 2017). However, the recent availability of new whole-genome assemblies based on long-read sequencing gives us the unprecedented opportunity to characterize complex forms of sequence variation that may have previously been overlooked (Rech et al 2021, Chakraborty et al 2019). This is of particular importance with regards to transposable element insertions, which are often associated with changes in gene expression under stressful conditions (Horvath *et al,* 2019,. Schmidt and Robin 2011a; Guio *et al.* 2014; Mateo et al 2014; Merenciano et al 2016; Ullastres et al 2019). Indeed, a natural transposable element insertion in the MTF-1 targeted gene *kuzbanian* has been associated with increased tolerance to zinc in adult flies, although the effect of the insertion was background dependent (Le Manh et al 2017).

In this study, we set out to assess variation in copper tolerance between natural populations of European *D. melanogaster* and investigate whether or not the phenotype is influenced by either geographic factors, the concentration of copper in soils, atmospheric pollution or degree of urbanization. To better elucidate the genetic basis of copper tolerance in natural populations, we compared the transcriptomes of three copper tolerant and three sensitive lines from before and after copper treatment, using a combination of tissue enrichment analysis, gene ontology, and modular clustering, to examine patterns of gene co-regulation. Finally, we also investigated the physiological traits relevant for copper tolerance. We found that while copper tolerance is highly variable across much of Western Europe, the external factors involved in shaping these phenotypes are complex, likely controlled by multiple regulatory factors, and that tolerance is linked to gut physiology.

RESULTS

Copper Tolerance is a Variable Trait Across European *D. melanogaster* Associated with Degree of Urbanization

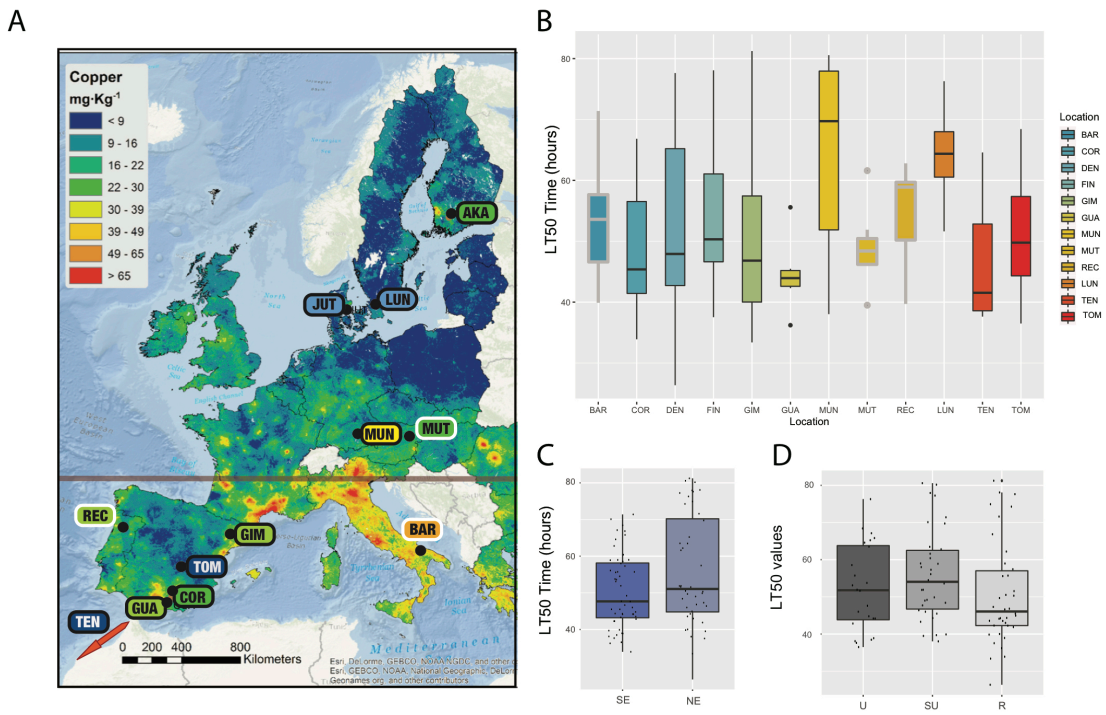
To assess the degree of copper tolerance in natural populations of *D. melanogaster* in Europe, we scored a total of 73 inbred lines taken from nine locations for copper mortality on a single dose over several days until full mortality was achieved (Fig. 1A, Appendix 1). LT_{50} values ranged from 26.4 to 81.2 hours, with a median value of 49.8 hours (Fig. S1A, TableS1). We observed minimal zero-dose control mortality during the course of the assay (Table S1). The LT_{50} values for all lines display a gradual, near-normal distribution, with a skew towards the sensitive end (Shapiro-Wilk W(70) = 0.934, p-value = 0.001). Although we observed a high degree of within-population variance in copper tolerance (Fig. 1B), pairwise comparison between the populations were not significant after the *Benjamini-Hochberg* correction was applied. (Table S2A). The near normality of this distribution suggests that the basis of copper tolerance at this given dose is heterogeneous, and most likely polygenic in nature (Battlay *et al.* 2016; Green *et al.* 2019).

As stress tolerance is frequently clinal in Drosophila (Hallas *et al.* 2002; Hoffmann and Weeks 2006; Arthur *et al.* 2008), we compared the differences in tolerance between northern and southern populations. We first performed a Wilcoxon's rank-sum test of the LT_{50} values between all samples as divided by the 45th parallel, as our nine collections sites could be clearly divided by this feature (Fig. 1A). The resulting p-value of $1.2x10^{-02}$ suggested that the differences in copper tolerance were significant. However, because all southern populations were collected in Spain, we broaden the analysis by phenotyping an additional 19 lines from Portugal and Italy in the south of Europe, along with another 7 lines from Austria (Fig. S1B). The inclusion of these new data did not strengthen the association between tolerance and geography, but rather weakened it, with a revised p-value of = 0.19 (Fig. 1C). We further examined the relationship between geography and copper tolerance by fitting generalised linear models between latitude, longitude and the LT_{50} values across all twelve locations. Longitude was considered in addition to latitude, because European *D. melanogaster* have been reported to exhibit population structure along this axis (Kapun *et al.* 2020). However, no significant correlations could be found (Table S2B).

As our initial interest in copper was spurred in part as its role as an environmental contaminate, we tested the relationship between copper tolerance and metal pollution. The measures considered — copper concentration in topsoils ($mgkg^{-1}$) and atmospheric pollution (PM10 and PM2.5; general and metal specific)—were based on what data was publicly available (see Material and Methods). Copper topsoil values varied widely across collection sites: from $6.7mg.kg^{-1}$ in Tenerife to $41.1\ mg.kg^{-1}$ in Bari (Appendix 1). However, despite the broad differences between locations, no general trend in tolerance could be observed after performing a linear regression between soil concentrations and LT_{50}

values by location (F = 2.272, p-value = 0.1626; Table S2C). Similarly, none of the atmospheric pollution measures available - either alone, or with soil copper added as an additional predictor variable – were associated with the degree of copper tolerance (Table S2C, Appendix 2).
As copper contamination is often the result of a complex group of contamination sources, especially around urban areas (Romic and Romic, 2003), we also considered a more indirect measure of pollution. We classified each of the catch locations into urban, semi-urban and rural classes, based on distance from highly populated areas (Appendix 2). Wilcoxon's rank-sum test of LT_{50} values across all three categories showed significantly lower LT_{50} values from rural locations when compared to those from semi-urban locations (Wilcoxon's rank sum test: urban vs. rural = 0.468; urban vs. semi-urban = 0.243, rural vs. semi-urban = 0.037; Fig. 1D).

Figure 1. Sampling locations and variation of copper tolerance across Europe
A) Distribution of the nine locations from the 2015 DrosEU collection (black border), and three additional collection locations (white border). Label fill corresponds with the copper concentration legend. The map shows the spatial variation in copper topsoil concentrations, as obtained from the Land Use/Land Cover Area Frame Survey (LUCAS) topsoil database, whose samples were taken from 2009 onwards. The 45th parallel is marked by a grey line. **B)** Boxplots of LT_{50} values by location. Locations from the 2015 DrosEU collection are outlined in black, while the three additional locations are outlined in grey. The full list of lines used are provided in Appendix 1. **C)** Boxplots of LT_{50} values of lines split into northern and southern locations by the 45th parallel. This point was chosen due to the fact the nine original collection sites could be clearly divided in half by this line. **D)** Boxplots of LT_{50} values of lines separated by degree of urbanization. In all cases, for each line, three replicas of 15 females for treatment and 10 females for control conditions were performed)



Tolerant and Sensitive Lines Demonstrate Differential Expression Profiles After Copper Exposure Mostly Concentrated in the Mid-gut

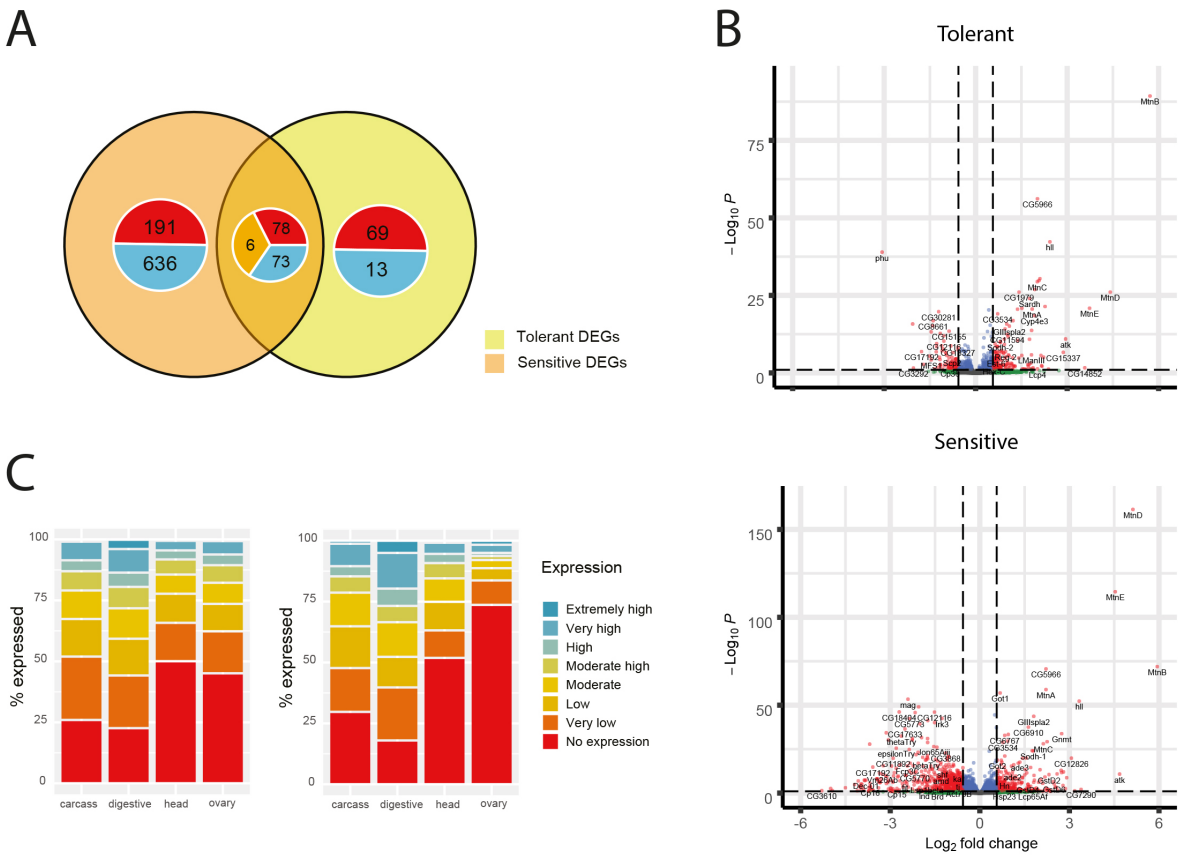
To examine the gene regulatory changes that occur in *D. melanogaster* in response to copper exposure, we compared the female whole-body transcriptomic profiles of three tolerant (GIM-012, MUN-020, MUN-008) and three sensitive lines (AKA-018, JUT-008 and COR-018), chosen primarily on the basis of their position at the tails of the phenotypic distribution (Fig S1a; see Material and Methods). Across the three tolerant lines, 239 genes were significantly differentially expressed (1 $\log_2$ fold change and adjusted p-value < 0.05) between copper treatment and control groups, while 984 genes were differentially expressed across the three sensitive lines, with an overlap of 152 genes (Fig. 2A). Of these 152 genes, the direction of the change was discordant in six, being all up-regulated in tolerant strains and down-regulated in sensitive (Table 1). The proportion of down-regulated genes was higher in the sensitive lines, with most of these down-regulated genes unique to sensitive lines (Fig. 2B and Table S3). These differences between tolerant and sensitive lines are also reflected in the Principal component (PC) analysis, where treated and controls samples from the JUT-008 and COR-018 showed a much greater separation on PC2 projection (Fig. S2A).

As expected for metal treatment, the metallothioneins *MtnA-MtnE* were the most significantly differentially expressed by a large margin (Fig. 2B). While there was no relationship between their degree of induction and tolerance, all six lines were found to carry the 3' indel polymorphism in *MtnA* that had previously been linked to oxidative stress resistance (Catalán *et al.* 2016). Other genes

previously documented to play a role in copper homeostasis were notably absent from the differential expression lists, including the *Crt1* family of transporters, *ATP7, Ccs* and *Malvolio.* While this is not necessarily an indication that genes involved in basic copper homeostasis are not involved, it does indicate that the biological basis behind copper tolerance in *D. melanogaster* may be constrained by the need to maintain copper levels in less extreme environments, and that increased tolerance goes beyond metal chelation and homeostasis.

In an attempt to find the tissues displaying the greatest levels of transcriptomic change, we used DGET to classify our DEGs — taken from whole body samples — according to their degree of expression in four of the tissue databases (head, carcass, gut and ovaries of four-day old females). (Fig. 2C). We focused on those genes with the higher levels of expression (those categorized as having either extremely high or high expression, with RPKM values greater than100).
We found that the greatest level of transcriptomic change was seen in transcripts from the digestive system (Tolerant p-value= 2.55e-19; Sensitive p-value = 9.59e-36; Fig.2C). The genes expressed in the gut included the five metallothioneins and a large number of serine peptidases, whose decrease in expression were especially notable across sensitive strains. Enrichment was also significant to a lesser degree in the carcass (Tolerant p-value= 4.58e-07; Sensitive p-value = 7.44e-14; Fig.2C). No DEG enrichment was seen for either the head or ovaries.
Regarding gut subsections, the most notable levels of enrichment were found in the copper cell region and the posterior gut (Fig. S2B). Copper cells are responsible for copper storage (Filshie *et al.* 1971; Tapp and Hockaday 1977), and changes in gut acidity (Marianes and Spradling 2013). One such marker of gut acidity —Vacuolar-type H+ATPase (*Vha100-4*) (Hung *et al.* 2020) — was found down-regulated by 0.6 (p= 0.01) and 2.0 (p=1.66E-08) across tolerant and sensitive strains respectively, suggesting that gut acidity may be playing an important physiological role.

Figure 2. Copper differential gene expression and tissue analysis
A) Venn diagrams showing the degree of overlap between the differentially expressed genes across the three tolerant and three sensitive lines. The numbers represented in red are found commonly up-regulated, those in blue commonly down-regulated and those listed in yellow are the genes with discordant changes in expression between tolerant and sensitive samples (Table 1). Expression data obtained from female whole-body RNA-seq (3 replicates of 20 females each for treated and control conditions). **B)** Volcano plots of gene expression in tolerant (top) and sensitive lines (bottom). The horizontal dashed line represents the minimum p-value threshold, while the vertical dashed lines represent the fold change thresholds. **C)** DGET expression analysis by tissues of the differentially expressed genes from the tolerant lines (left) and differentially expressed genes from the sensitive lines (right).



Metabolism, Reproduction and Protease Induction Contribute to Copper Response in Tolerant and Sensitive Lines to Different Degrees

To determine what biological and physiological processes might differ between the tolerant and sensitive strains after the same period of copper exposure, we performed gene ontology (GO) enrichment analysis on each group (Table S4). Metabolism related terms were commonly seen as the largest and most significantly overrepresented terms in both tolerant and sensitive lines, although the exact processes shift between the two groups often varied (Fig. 3A). Chitin metabolic process (GO:0006030, tol p=3.41E-6; sen=p9.23E-13) and Chitin binding (GO:0008061, tol p=8.99E-6,; sen=p1.1E-13) were also common to both analyses (Table S4). As expected, response to metal ion (GO:0010038) was strongly overrepresented in the three copper tolerant lines (tol p=4.10E-6, Fig. 3A). Additionally, several GO terms linked to reproduction and vitellogenesis—including chorion-containing eggshell formation (GO:0007304; sen p-value = 1.66E-9,) and loss of vitellogen (encompassed by GO:0030704; sen p-value = 6.36E-08,)— were found to be overrepresented in both analysis, but more so in sensitive lines (Fig. 3A and Table S4). Note that shutdown of egg production is often a consequence of heavy metal toxicity (Terashima and Bownes 2005; Ojima et al. 2018). Increasing the fold change cut off for GO enrichment to 1.5 reduced the significance of the metabolic based terms while further highlighting the differences in metal ion response in the tolerant lines, and

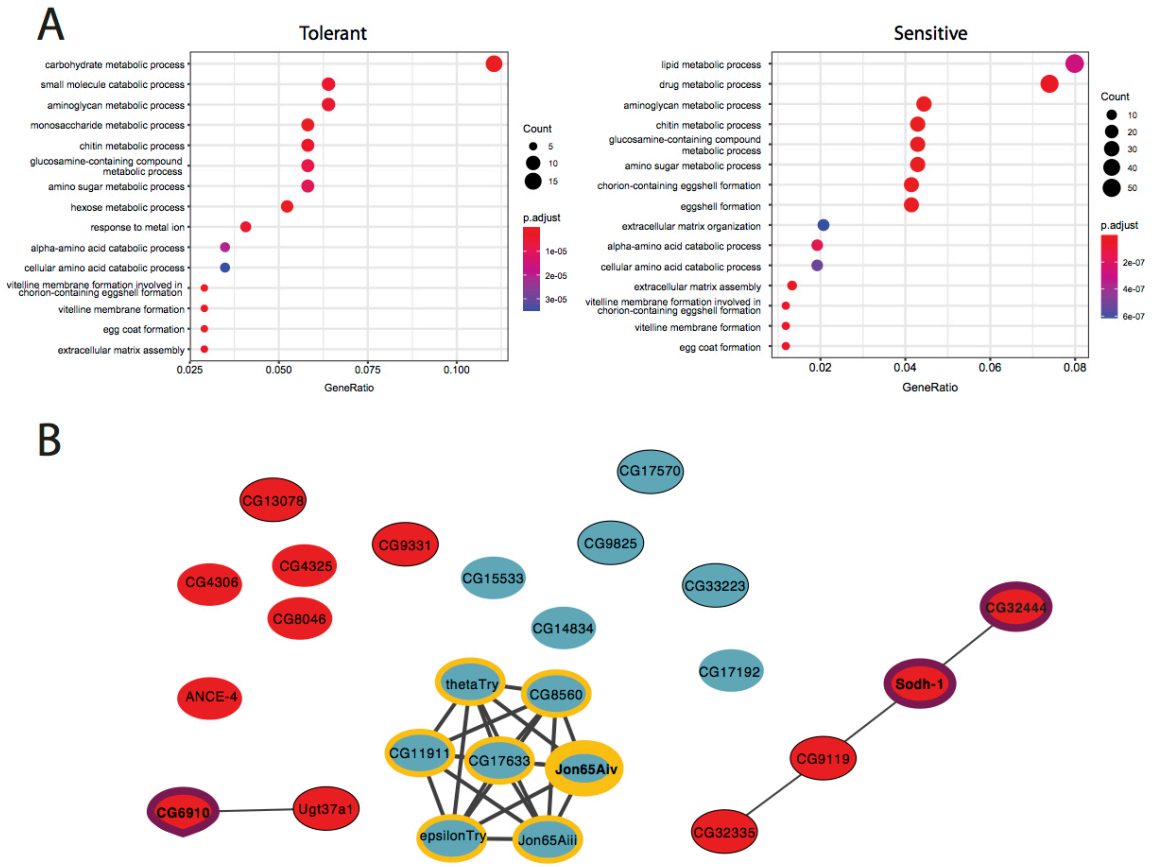
reproductive response in the sensitive lines (Fig. S3). The majority of these terms were also found to be overrepresented in Gene Set Enrichment Analysis (GSEA) (Table S5). Further KEGG analyses also emphasised the role of protein metabolism processes in copper stress response and lysosome activity both in tolerant and sensitive strains (Table S6). Overall, these results suggests that under our assay conditions, the more tolerant lines are undergoing metabolic stress response at the end of the assay, while the more sensitive lines had progressed to shutting down non-essential biological processes—such as egg laying—at the same stage

We also investigated the level of gene co-regulation in response to copper of tolerant and sensitive strains using modulated modularity clustering (MMC) analysis, which in contrast to previous analyses does not rely on any prior gene functional annotations (Stone et al 2009). Tolerant lines show a higher level of expression co-ordination after copper exposure while sensitive lines showed the opposite pattern (Fig. S4 and Table S7). Heat-maps of the tolerant treated lines suggested that genes in modules 2-5 were very closely linked (Fig. S4, Table S7). Analysis of the 25 genes represented by these modules in STRING (Szklarczyk *et al.* 2018) revealed a group of seven tightly interacting serine peptidases (Fig 3B) that are found highly expressed in the digestive system. While a number of these genes were from the *Jonah* family of serine peptidases, the discordantly expressed *Jon99Ci* was not included among them (Table 1). Of these seven serine proteases, four have previously been shown to be regulated by the histone and protein deacetylase *Sirtin 2* (*Sir2;* Palu and Thummel 2016). On further inspection, 58candidate genes from our tolerant strains and 187 from our sensitive strains have previously shown to display differential expression after *Sir2* knockdown, a significant overlap (Hypergeometric test: Tolerant = 1.30e-20; Sensitive = 6.87e-45; Fig. S5A) (Palu and Thummel 2016). *Sir2* has multiple downstream targets – *DHR96, dfoxo* and *HNF4* – which have been shown by Palu and Thummel (2016) to alter the expression of overlapping sets of genes. While a small degree of overlap was seen between our differential expression lists and that of *dfoxo* (Hypergeometric test Tolerant = 3.10e-07; Sensitive = 1.50e-20) and *DHR96* knock out analyses (Hypergeometric test Tolerant = 3.04e-07 Sensitive = 3.58e-12. ; Fig. S7), the greatest overlap was seen with *HNF4* knock out analyses (Hypergeometric test Tolerant = 1.45e-19; Sensitive = 8.94e-66; Fig. S5B) (King-Jones *et al.* 2006; Alic *et al.* 2011; Barry and Thummel 2016). While little is known about the precise role *HNF4* plays in the midgut, its inferred link to the serine peptidases suggests a potential role in digestion.

Figure 3. GO and correlational clustering analysis

A) Top 15 enriched GO terms associated with the DEGs in tolerant and sensitive strains. The Y-axis indicates gene functions, and the X-axis indicates the percentage of total DEGs in a given GO category (gene ratio). **B)** String interaction network of candidate genes taken from highly correlated

modules (2 to 5) of the MCC analysis for the treated samples of the tolerant strains (PPI enrichment p-value: < 1.0e-16.) Upregulated genes are shown in red, down-regulated in blue. The genes highlighted in yellow are part of the serine peptidase cluster, while those with borders in bold were later selected for further validation using RNAi knockdown or gene disruption lines.

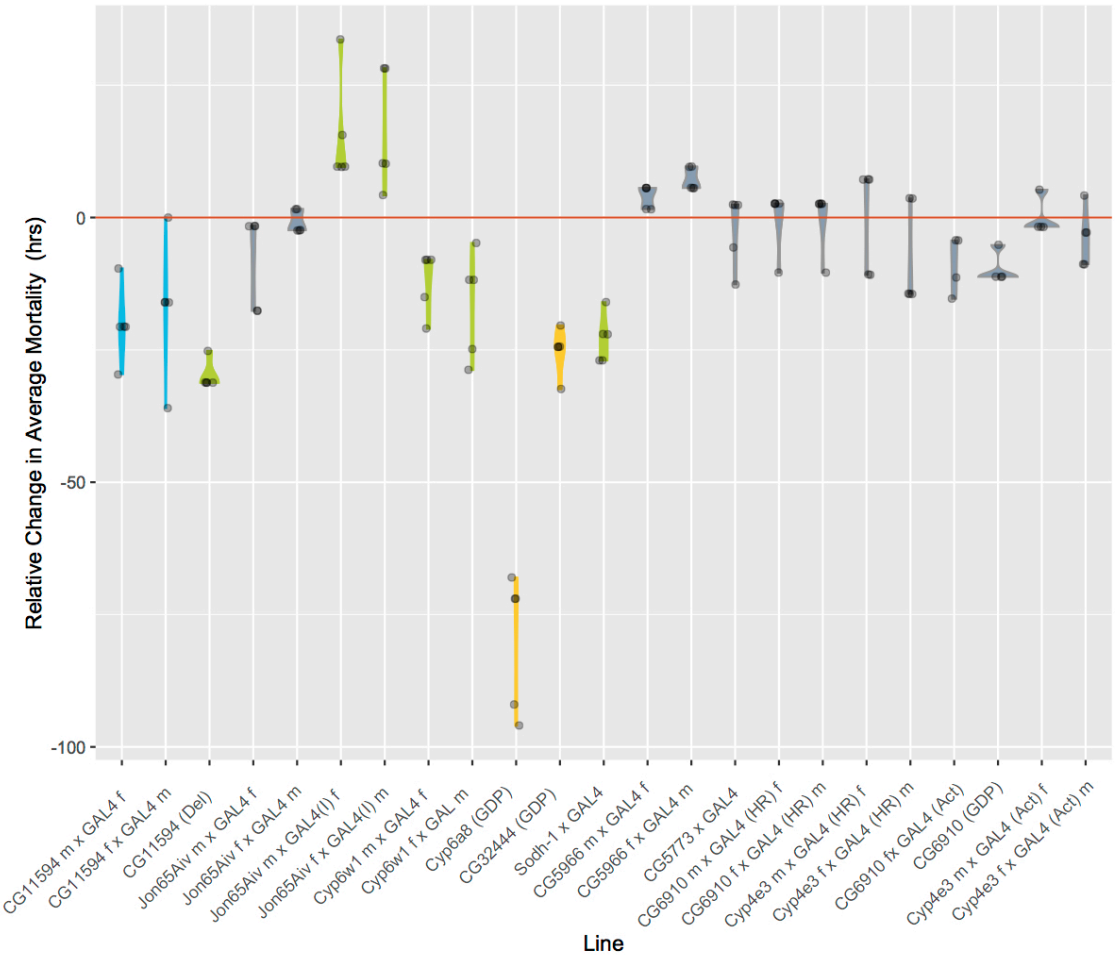


Seven out of 10 Copper Candidate Genes Were Confirmed to Play a Role in Copper Tolerance. Ten of the candidate genes associated with copper response based on our transcriptomic analysis were chosen for further characterisation. Three of these genes — *CG5773*, *CG5966* and *Cyp4e3* — were chosen on the basis of their differential expression data alone. Three other candidates— *CG6910*, *CG32444* and *Sodh-1*—have been linked to copper homeostasis previously in the literature, but their exact functions have not been well characterised (Schmidt *et al.* 2000; Southon *et al.* 2004; Yepiskoposyan *et al.* 2006). The remaining four candidate genes, *CG11594*, *Jon65Aiv, Cyp6w1* and *Cyp6a8* were all found to be associated with TE insertions (see below). In addition, four of the ten candidates were part of the MMC cluster containing the serine peptidases (*Jonah65Avi*, *CG6910*, *Sodh-1* and *CG32444*; Fig 3B).

Six of the genes tested showed changes in phenotype when knocked-down or disrupted in the direction expected based on our expression data (Fig. 4, Fig S6, Table S8). Of the six confirmed genes

CG11594, *Jon65Aiv*, *Cyp6w1* and *Cyp6a8* are novel candidates, whose full role in copper biology is not yet understood. *CG32444*, and *Sodh-1*, which have some prior links to the phenotype (Everman *et al.* 2020; Southorn *et al.* 2004), showed decreased survivorship when knocked-down as well. However, more unusually, *CG5966* displayed increased survivorship when knocked-down, which was not predicted by its induction on copper (Fig. 4, Fig S6, Table S8).
CG5773, *CG6910* and *Cyp4e3* did not display any changes in survivorship after knock-down (Fig. 4 and Table S8). As both *CG6910* and *Cyp4e3,* were initially tested with *6g1HR-GAL4* driver, based on prior expression data (Leader *et al.* 2017), we repeated the crosses with the *Actin5C-GAL4* driver, along with screening a Gene Disruption Project (GDP) insertion line for *CG6910.* However, none of these three additional assays showed any significant changes in survival. While it is possible that the effects of these genes on copper tolerance are background sensitive, as per the example of *Cyp6g1* (Deneke *et al* 2017) and *Cyp12d1* (Green et al, 2019), it is also possible that these genes have little to no true impact on the phenotype at all, and are only present due to co-regulation with other genes that do directly affect copper tolerance—a phenomenon that has been observed with regards to the *Cnc*/*Keap1* pathway (Kalsi and Palli 2017) (Fig. 4, Fig S6, Table S8).

Figure 4. Relative copper survival for all ten copper candidate genes.
Relative change in average mortality at the end of the assay comparing candidate gene disruption and knock-down with their controls (3 to 5 replicates of 15 females for the treatment and 10 females for the control conditions). Significant relative mortality with a p-value <0.05 is shown in blue, in green relative mortality with a p-value <0.05-0.01, and yellow shows p-values < 0.01. Del is deletion and GDP is gene disruption. For RNAi knockdowns, genes thought to act in the 'detox' tissues—including the gut fat body and malphigian tubules—were targeted with the *6g1*HR-GAL4 driver, while those genes whose expression was more gut specific were targeted with MexG-Gal4. An ubiquitous Actin5C-GAL4 knockdown was used for all other crosses. Drivers abbreviations are as follows: *GAL4(I)* is the introgressed version of GAL4; *HR* is the *HikoneR* driver and *Act* is the *Act5C* driver.



Copper Tolerance is Correlated with Gut Acidity and *CG11594* Activity and not Mitigated by Changes in Feeding Behavior.

Both our DGET analysis and our GO analysis have lent evidence to the idea that changes to the gut could be linked to copper tolerance (Fig. 2C and Fig. 3A). As copper accumulation in *D. melanogaster* has previously been linked to changes in gut physiology (Cioffi 1984), we assayed the changes in gut pH after copper exposure. Adults from the six RNA-sequenced lines were subject to copper assay conditions, and then allowed to recover for two hours on regular media supplemented with a mixture of Bromophenol Blue and yeast. If the acidic copper cell region of the gut remains un-inhibited by copper, this region should remain yellow under Bromophenol Blue yellow (pH<2.3). After two hours, most recovering individuals had consumed enough media for the dye to be detected in the gut. Only AKA-018 had more than 10% of flies failing to feed on recovery– a phenomenon that was not seen in the controls. While all six lines showed decreased acidity across the copper cell region after comparing treated and control conditions, the three sensitive lines showed a much higher loss of acidity than the three tolerant (Fig. 5A and Table S9A). Greater than 25% of individuals across all

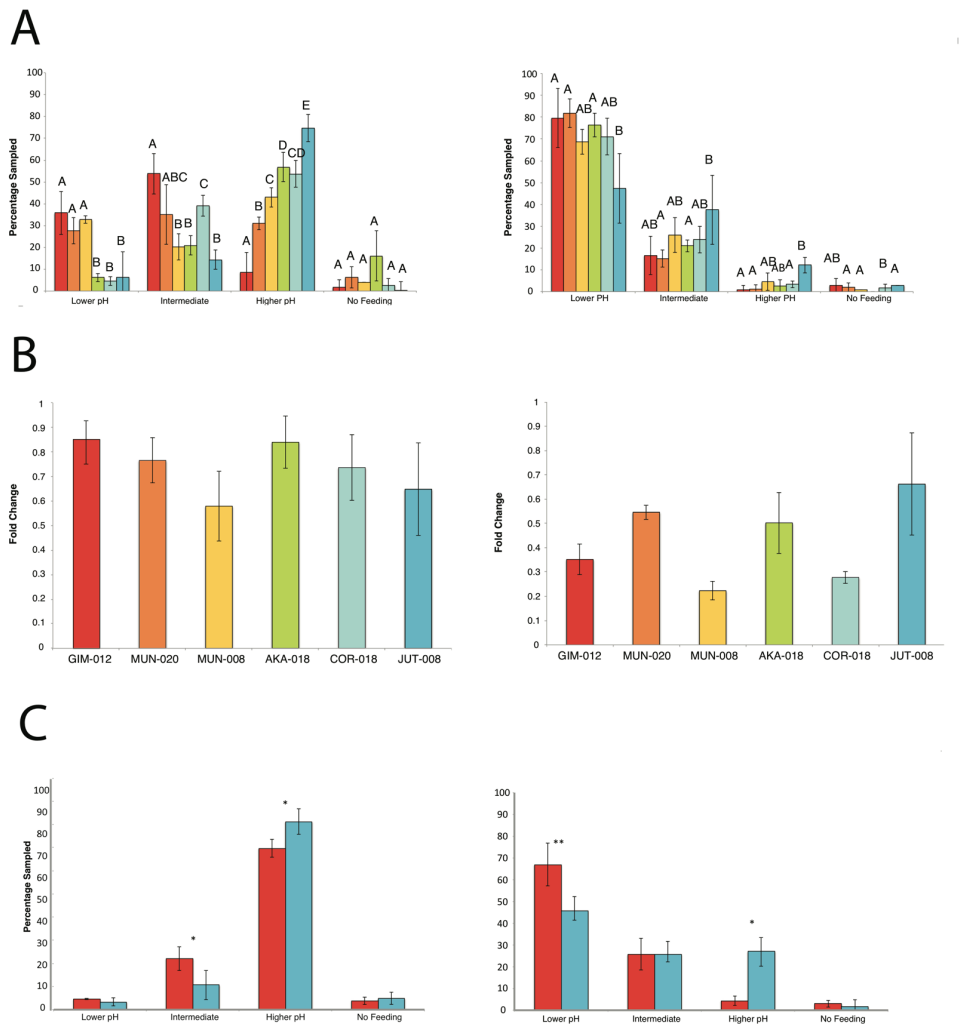
tolerant lines maintained a low pH under treated conditions compared to less than 10% of the sensitive lines (χ^2 p-value = 4.815e-07). Differences were less pronounced under control conditions, with only JUT-008 showing an appreciable loss of acidity in the absence of copper. From these results, we can infer a link between the loss of acidity in the copper cells and a decreased ability to tolerate copper.

As previously noted by Bonilla-Ramirez *et al,* (2011), *D. melanogaster* often avoid food sources with high concentrations of heavy metals. To determine if the changes in gut acidity are influenced by changes in feeding behavior, we repeated the copper tolerance assays on the six RNA-sequenced lines, this time with the addition of a 1% Erioglaucine Disodium Salt to both the treatment and control solutions to act as a dye. We measured the level of dye consumed in both treatment and control conditions at two separate time-points to determine the level of feeding avoidance. At 24 hours, the level of feeding avoidance on copper across most of the lines was quite low when compared to their control counterparts, with no significant differences between lines (Fig. 5B and Table S9B). Feeding avoidance was generally stronger at the 40-hour mark, where both MUN-008 and COR-018 were distinguishable by their greater levels of feeding avoidance (Fig. 5B and Table S9B). However, no relationship could be drawn between feeding behavior and whether or not the line showed high or low copper tolerance or changes in gut acidity.

If the observed changes in gut acidity are not based in behavior, they are likely physiological in nature. Of the seven candidate genes we confirmed as having a role in copper tolerance, *CG11594* is the most poorly characterized of the candidate genes with no prior links to copper biology. To determine if *CG11594* expression alters gut acidity, similar exposure and gut staining assays were carried out over a sixteen hour time periods on a *CG11594* deletion strain, with *w[1118]* as the control strain. While both lines displayed a high degree of gut de-acidification after treatment than any of the six natural lines, the effects seen on the *CG11594* deletion line were significantly greater than those on the control line (Fig. 5C and Table S9C).
Curiously, the clearest differences between the two lines were seen not in the treatment arm, but in the control, where only a half of the *CG11594* deletion individuals displayed a clearly defined acidic region. This is in stark comparison to the six sequenced lines, which displayed healthy guts under control conditions. These results indicate that physiology, not behavior, is the main driver behind midgut de-acidification after copper exposure, and that *G11594* plays a role in this change.

Figure 5. Gut acidity after copper exposure is correlated with copper tolerance and is not linked to feeding behaviour

A) Gut acidity results on the six RNA-sequenced lines after 24 hours of copper treatment and 2 hours of recovery (left) and after 24 hours of control conditions and 2 hours of recovery (right). Lower pH indicates that the dye turned yellow within the region of the gut containing copper cells; intermediate pH indicates that the dye turned green-brown, but a discrete acidic region could still be detected; higher pH indicates that the entire midgut was blue and the copper cell region could not be detected; and no feeding (clear or pale blue). **B)** Feeding avoidance in the presence of copper measured as a fold difference in consumption between treatment and control at 24 hours (left) and 40 hours (right).. **C)** Gut pH of the control strain *w[1118]* (in red) and the *CG11594* deletion strain (in blue) after 16 hours of copper treatment (left) and control conditions (right). An asterisk (*) indicate a difference across the treatment groups at $p < 0.05$ (*) and (**) $p < 0.01$. For all the plots, error bars represent the standard error of the mean of three replicates containing 30-45 females each (A and C) and three replicates containing 25-30 females each (B).



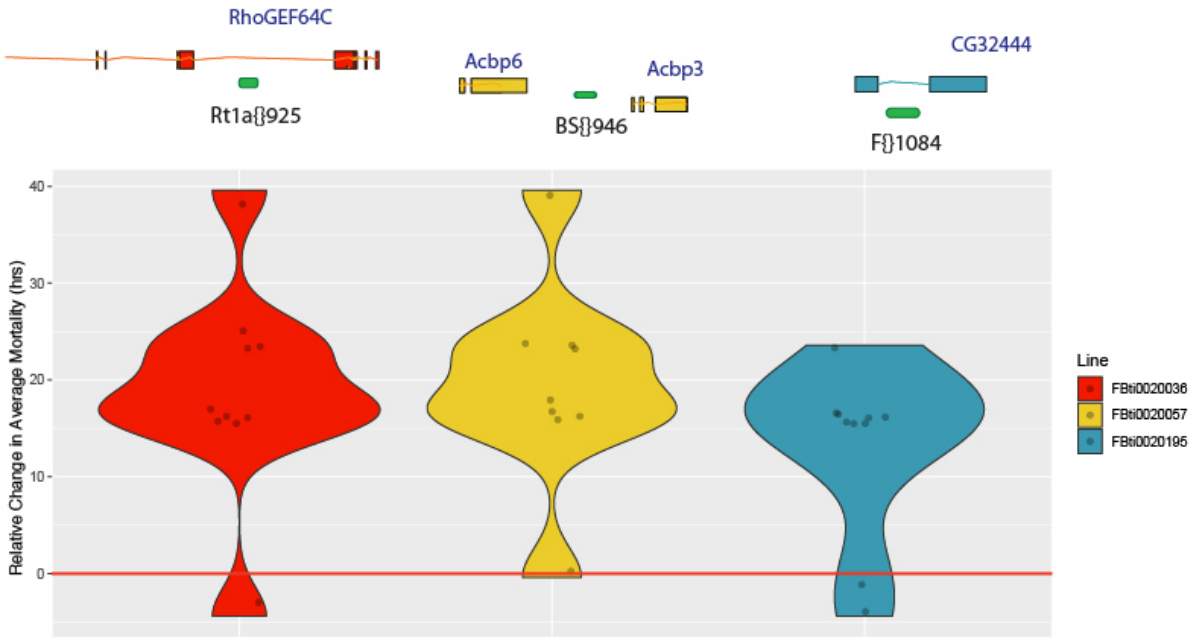
Transposable Element Insertions May Influence Copper Tolerance.

TE insertions are often associated with changes in gene expression under stressful conditions (Horváth *et al.* 2017), and in *D. melanogaster*, several specific insertions have been linked to stress response including zinc stress (*e.g.* Schmidt and Robin 2011a; Guio *et al.* 2014; Mateo et al 2014; Merenciano et al 2016; Le Manh et al 2017; Ullastres et al 2019). However, until recently only the subset of TEs annotated in the reference genome could be analyzed limiting the power of genome-wide analysis for this type of structural variants. We took advantage of the availability of *de novo* whole genome assemblies and *de novo* TE annotations for the three tolerant and three sensitive lines analyzed in this work to investigate the association between proximal *cis* TE insertions and gene expression levels in both treated and control conditions. (within 1kb of the insertion, see Material and Methods). Using QTLtools, we identified three TE insertions that were significantly associated with changes of expression in nearby genes: two in response to copper and one both in control and in response to copper (Table S10A). Although the number of significant associations is small, this is most probably due to the small number of genomes analyzed (six)— suggesting that this approach should provide more insight with larger datasets. As an alternative approach, we also investigated whether previously identified DEGs in tolerant and sensitive strains were located within 1kb of a TE insertion (Table S10B). We found that the percentage of differentially expressed genes located within 1kb of a transposable element was higher in tolerant compared to sensitive lines, although the difference was not significant (14.28% across the three tolerant lines and 11.29% in the three sensitive; p-value= 0.2193). No particular TE location, either upstream, downstream or within the gene, was favoured. While 73.5% of the TE insertion were associated with gene up-regulation in tolerant strains only 28% of the TEs were associated with up-regulation in sensitive strains (Fisher exact test p-value = 0.0014; Table S10B).

Finally, we tested three TE insertions for their effects on copper tolerance. To do this, we chose two insertions that besides being located nearby DEGs, showed signatures of selection in their flanking regions, suggesting that they might be adaptive: *FBti0020036* and *FBti0020057* (Rech *et al* 2019). The third TE candidate, *FBti0020195,* is not present in any of our six sequenced lines, but garnered special interest due to its location within *CG32444,* a candidate gene we had previously confirmed with the use of gene disruption (Fig. 6). For each of these three TE insertions, we constructed two outbred populations: one with the insertion and one without the insertion (see Material and Methods). For each of the paired outbred populations, those containing TE insertions all demonstrated greater survivorship on copper than their negative counterparts (Fig. 6 and Table S10C). There is no clear relationship between the direction of the changes of expression seen in the gene candidates physically closest to the TEs and the increase in copper tolerance, suggesting that the insertions are not having the same effects at each locus.

Figure 6. Relative copper survival for the three candidate transposable elements

Relative change in average mortality at the end of the assay comparing outbred populations with and without the candidate TE (9 to 10 replicates of 15 females in treatment and 10 females in control conditions). On top, gene structure showing where the candidate TE is inserted. For *RhoGEF64C* only the 3' region of the gene is depicted.



DISCUSSION

The Environmental Determinants of Copper Tolerance in *D. melanogaster* are Complex

In this study, we undertook a survey of multiple European *D. melanogaster* populations to determine how copper tolerance varies across the continent, and whether this variation could be linked to the presence of copper or other environmental factors. To achieve this, we compared our phenotypic values with geographic factors, copper soil levels, atmospheric pollution levels, and degree of urbanization. While we found evidence of a link between urban build-up and greater tolerance, not clear relationship could be drawn between tolerance and any of the direct measures of pollution available to us. As Romic and Romic (2003) noted, human sources of environmental copper are characterised by many point sources of contamination, and while we are aware that some well-known sources — such as atmospheric copper — are missing from our dataset, it is possible that there are others missing as well. It is also unknown whether the greatest effect will be from an accumulation of multiple sources of the metal, or a small number that are the most bio-available. Our categorical analysis between urban, semi-urban and rural locations is suggestive of the latter, as the significant

differences were found between semi-urban and rural categories (Fig. 1D). As these point sources can be difficult to characterise, performing environmental sampling, *e.g.* soil sampling, alongside fly collections may be a viable alternative (Massadeh et al 2008). The diversity of vegetation may also be worthy of record as copper uptake and storage varies across plant tissues and species (Adriano 2013). Although we cannot discard that more extensive sampling could further help discern the relationships between phenotype and environment, our results indicate that the finer details of the surrounding environment should be receiving as much attention as the finer details of the genome when making sense of phenotypic differences.

The Genetic Basis to Copper Tolerance in *D. melanogaster* is Complex, and Involves Multiple Regulatory Factors

One of the most distinguishing features of our phenotypic dataset is the high degree of variation both within and between sampling locations (Fig1B and Fig S1). While high levels of phenotypic variation can sometimes result from an allele of large effect segregating within a population, as seen in Battlay *et al.* (2016) and Green *et al.* (2019); the gradual distribution of our LT_{50} values suggest that this is not the case and that the degree of phenotypic variation seen across our lines is likely an indication of the polygenic basis of the trait. This was in turn backed by our RNA-sequencing analysis, which indicated that copper tolerance is a trait with a complex genetic architecture; involving multiple genes and regulatory factors, and with a large degree of expression change occurring in the gut.

With regards to genes with prior links to metal response, variation in metallothionein expression was not found linked to phenotypic variation in the six lines sequenced (Fig. 2B). However, as all six lines carry the 3' indel believed to be close to fixation in northern Europe (Catalan et al 2018), thus it is likely that tolerant variants between the metallothioneins have already been subject to selection. We also saw no significant differences in expression with regards to multiple genes previously linked to copper homeostasis; leading credence to the idea that the genes or alleles that affect variation in copper tolerance in natural populations are different to those in genetically modified lines. While this may initially come across as curious, many of the previous studies characterising copper-related genes in genetically modified lines were carried out in a small number of strains with the aim of characterizing genes that play a role in human diseases (Southon *et al.* 2004; Norgate *et al.* 2007; Southon *et al.* 2008), and not over copper toxicity. While the lack of these genes in our DEG lists does not necessarily mean that they have no involvement in copper tolerance, it does indicate that the genes contributing to the variation we see in tolerance in natural populations of *D. melanogaster* are much broader than previously characterised in these studies, and that increased tolerance goes beyond metal chelation and homeostasis.

While it has been well documented that MTF-1 plays an important role in regulating gene expression in response to metal exposure, including metallothionein induction, it is unlikely to be the only regulatory factor affecting changes in expression, especially with regards to downstream metabolic processes affected by copper toxicity (Roelofs et al 2009). By using a combination of DGET and gene clustering, we were able to identify *Sir2* and *HNF4* as additional potential regulatory elements. *Sir2* plays a multifaceted role in maintaining energy homeostasis, affecting fat mobilization (Banerjee et al. 2012), insulin signalling (Banerjee et al. 2017), and energy consumption (Palu and Thummel 2016). *HNF4*—a direct target of *Sir2* regulation—also influences a wide range of processes involved in cellular metabolism and systemic physiology (Barry and Thummel 2016). These results are supported by our functional gene analysis. Of the seven confirmed candidate genes (Fig. 4), *Sodh-1* and *CG32444* have both been linked to the kind of metabolic processes modulated by *HNF4* and *Sir2*, while also having found associated with copper toxicity previously (Everman et al 2021, Southon et al 2004). The two Cytochrome P450s, *Cyp6w1* and *Cyp6a8*, are both linked to oxidative stress (Misra et al 2011), a process that has been linked to metal tolerance previously in both insects (Roelofs et al 2009).
The roles of the remaining three candidates in copper tolerance are more speculative. While *Jon65Aiv* is known to be a serine peptidase with a likely role in digestion (Ross et al 2003), and *CG5966* is involved in triglyceride breakdown (Wat et al 2020) and regulated by both *Sir2* and *HNF4*, our pH assays give the greatest guidance to the role of *CG11594*, which appears to play a role in gut integrity.

While our individual gene candidates may not be so well conserved outside of Drosophila, *Sir2* and *HNF4* do have well-conserved orthologs, much in the same manner as the metallothioneins. While there is no previous evidence for these genes playing a role in copper toxicity in arthropods, such evidence exists in mammalian cell culture: rat hepatocytes treated with copper sulphate display increased expression of *Sir2* homologues *Sirt1* and *Sirt2* (Sun et al. 2014), while *HNF4-α*, influences copper responsive transcription changes in HepG2 cells (Song and Freedman 2011). Furthermore, while many of our punitive candidates for *Sir2 and HNF4* regulation were found highly expressed in the gut, both regulatory elements have been shown to play different roles in different tissues (Barry and Thummel 2016; Palu and Thummel 2016), presenting us with the possibility that not only might their roles in copper response be discordant in different tissues, but that this may apply to the general transcriptional signature post metal exposure as well. Future assays using knockdown or disruption of these factors across multiple tissues in Drosophila would be able to confirm their specific roles in copper response.

Further changes in gene expression can potentially be traced back to transposable element insertions. TE insertions are often associated with the differential expression of nearby genes under stress conditions (Schmidt and Robin 2011b; Guio *et al.* 2014; Horvath *et al* 2017). We identify several TE

insertions located inside or nearby differentially expressed genes (Table S10B). For three of these insertions we further showed that their presence in associated with increased survival (Fig. 6). Further analysis, such as recombination mapping and CRISPR based knock-outs in these genetic backgrounds could potentially assist in confirming the role of these specific TE insertions in altering gene expression and their effect on phenotype.

Gut acidity is linked to copper tolerance in *D. melanogaster.*

Our analysis demonstrated that a large degree of the differential expression observed after copper exposure was occurring in the gut, a key tissue when it comes to copper physiology (Cioffi 1984, McNulty *et al.* 2001, Dubreuil 2004, Li *et al.* 2016). A role for the gut is also supported by the GO enrichment results: chitin binding and metabolic processes suggest a role for the peritrophic membrane (Kuraishi *et al.* 2011), which is important for gut integrity. A study on the effects of Lufenuron— a chitin disrupter— in *Anthonomus grandis* showed that gut disruption could lead to changes in metabolism and the down regulation of vitellogen; also seen in our GO enrichment analysis (Cruz *et al.* 2021). In addition, chitin binding and metabolic processes also affect the cuticle, which may affect copper exposure via contact. Indeed, and although less significant than in the gut, we also found DEGs to be enriched for extremely high and highly expressed genes in the carcass (Figure 2C).

Our gut pH assays clearly demonstrate that copper exposure results in a loss of acidity in the copper cell region—and that this effect is more sharply seen in the three sensitive lines (Fig. 5A). Our subsequent feeding response assays excluded differences in copper consumption as a potential explanation of varying losses in gut acidity, suggesting a more physiological process was responsible for the changes observed (Fig 5B). While metallothioneins would be obvious candidates, our *Mtn* expression data does not sufficiently explain the differences we observed (Fig. 2B). This opens up the possibility that one or more of our gene candidates selected for further analysis may be affecting copper tolerance through changes in copper cells or gut acidity. While the function of *CG11594* has mostly gone uncharacterised, it's expression has been linked to both oxidative stress and ER stress in the DGRP (Weber *et al.* 2012; Chow *et al.* 2013). While disruption of *CG11594* expression caused a strong loss in gut acidity after copper treatment, there was a notable loss under control conditions as well (Figure 5C). These results imply that loss of gut acidity is a sub-phenotype to copper tolerance, and that both share links to *CG11594* activity—although the exact mechanism underpinning the relationship remains elusive. In light of previous studies, we can propose two tentative alternative hypothesis: regulation of *CG11594* by both *Sir2* and *HNF4,* suggests that the gene plays a general role in energy and metabolism (Palu and Thummel 2016), and it is differences in the allocation of energy and resources that affects survival. Alternatively, links to ER stress (Chow *et al.* 2013) could indicate a role linked to lysosome function or metal storage.

Overall, our investigation across European natural populations of *D. melanogaster* proved copper tolerance to be a highly variable trait. We confirmed the involvement of multiple new candidate genes, identified two potential new regulatory factors that have previously only been seen to mediate metal responses in mammals, and described physiological changes linked to this trait. Unlike previous candidates, such as the metallothioneines, which are common across a wide phylogeny, it is unlikely that the exact poly-genes shown to affect copper tolerance in *D. melanogaster* will be perturbed in other species vulnerable to metal toxicity. However, other, more general, molecular pathways and physiological changes in the gut we observed in *D. melanogaster* are likely to prove relevant in studying the effects of copper toxicity in other species.

Material and Methods

Fly Collections

Details of all the stocks used can be found in Appendix 1. The nine original collections were carried out across the summer of 2015 by the DrosEU consortium. Each of the established isofemale lines was repeatedly inbred for up to 20 generations. Of the additional lines included for geographical and environmental analysis, the Mauternbach and Recarei lines were caught in 2018 and the Bari lines were caught in 2011 and have been kept as isofemale lines since then (Mateo *et al* 2018). All catch sites are documented in Figure 1a. All lines were maintained on semolina-yeast-agar media and were kept at 25 °C on a 12:12 hour light and dark cycle for at least one generation before use.

Copper tolerance assays

Copper Sulphate ($CuSO_4$) (CAT# 451657-10G) was obtained from Sigma Aldrich. Copper assays were adapted from Bonilla-Ramirez *et al.* (2011). This particular method was chosen for two reasons: (i) as Drosophila are known to show food avoidance with a high concentration of heavy metals (Bonilla-Ramirez *et al*. 2011) it allowed exposure via both contact and digestion; and (ii) the 4-5 day length of the assay gives sufficient time to differentiate between tolerant and sensitive lines without risking high control mortality.

Briefly, powdered $CuSO_4$ was reconstituted to 20mM in a 5% sucrose solution. Food dye was added to aid visibility and even dispersal. An identical control solution without $CuSO_4$ was prepared in the same manner. 250µl of the CuSO4 sucrose solution was pipetted onto 70x17mm slips of filter paper (Whatman, CAT# 3030917), which were then placed into 15ml Falcon tubes (Cultek, CAT# 352096), containing 1ml on 1% agar at the bottom. Papers were allowed to dry for 15 minutes before the flies were added. To assist respiration, holes were made in the lids of the falcon tubes.

4-7 day old females were used for all copper survivorship assays and 4-5 day old females for pH assays (see below). For assays on the inbred and isofemale strains, three replicates of 15 flies were performed for the treatment assay, and three replicates of 10 were performed for each control. For assays involving RNAi knockdown or gene disruption, three to five replicates of 15 were performed for the treatment, and five replicates of 10 were performed for each control. For transposable element presence and absence comparisons, this was increased to 9-10 repeats for both treatment and dose.

LT_{50} calculations were used to interpolate measures of survival for each of the lines. Linear models were fitted to time point-mortality data on a log-probit scale using the "glm" function in the R statistical package, using a script adapted from Johnson *et al.* (2013). Of the 73 DrosEU lines screened, LT_{50} values were successfully calculated for 71, along with the 26 additional lines from Italy, Austria and Portugal (Appendix 1).

Correlation Analysis with Geographical and Environmental Variables

Copper soil concentration data was taken from The European Soil Data Centre (ESDAC: https://esdac.jrc.ec.europa.eu/content/copper-distribution-topsoils), with the exception of the Tenerife data, which was taken from Fernandez-Falcon *et al.* (1994). Air pollution data was taken from the European Environment Agency (EEA): http://aidef.apps.eea.europa.eu. The pollutants considered included PM10 (particulate matter 10 micrometers or less in diameter), PM2.5 (particulate matter 2.5 micrometers or less in diameter); along with separate arsenic, cadmium and lead measures taken from the PM10 data. All measures were taken from the closest research station available for each catch site. General PM10 and PM2.5 data and atmospheric metal data for arsenic, cadmium and lead were available for the majority of catch sites (Appendix 2). Data for particulate copper taken from PM10 measures had to be excluded due to both insufficient geographical coverage and a lack of consistency in the measures made (PM10 and precipitation). All tests and linear regression models were performed in *R v3.5.1*. Regression models were fitted with LT_{50} values as the dependent variable, and with geographical and pollution measures as independent variables. Urban classification of catch locations were based on whether or not a catch site was within a city (urban), within 5km of a town with a population > 10,000 (semi-urban) or more remote regions (rural) (Appendix 2).

RNA-seq Sample Preparation

RNA-seq analysis for short-term copper exposure was performed on six inbred lines, of which GIM-012, MUN-020 and MUN-008 were copper tolerant and JUT-008, COR-018 and AKA-018 were copper sensitive (Appendix 1). To maximize odds of choosing mostly homozygous lines, we prioritized those lines with a high degree of inbreeding (minimum of F20), and a low degree of variation between biological replicates in the LT_{50} assays. Four biological replicates of 25 4 to 7 day-old female flies from each line were exposed to $CuSO_4$ or the equivalent control conditions from the

previous copper tolerance assays and removed after 24 hours - a timeframe that allowed low levels of death in the sensitive strains, but enough time to stress tolerant strains, as measured by the induction of *mtnB*. Deceased individuals from lines COR-018 and JUT-008 were removed. Total RNA was isolated from 20 females from each biological repeat using the GenElute Mammalian Genomic RNA miniprep kit (Sigma Aldrich, CAT# RTN350-1KT), following the manufacturer's instructions. For each sample, the three repeats with the best extractions were retained for sequencing. 1.5 µg of total RNA from each sample (whole female body) was used for subsequent library preparation and sequencing using an Illumina Hiseq 2500. Only two control samples for both AKA-018 and MUN-020 showed high enough quality for further analysis.

Analysis of RNA-seq Data

RNA-seq analysis was performed using the *rnaseq* pipeline v1.2 from the *nf-core* community, a *nextflow* collection of curated bioinformatic pipelines (Di Tommaso *et al.* 2017; Ewels *et al.* 2020). Briefly, sequencing quality was assessed using *FastQC v.0.11.8* (www.bioinformatics.babraham.ac.uk/projects/fastqc). *TrimGalore v.0.5.0* (www.bioinformatics.babraham.ac.uk/projects/trim_galore) was used for adapter removal, and *Cutadapt v. 1.18* with default parameters was used for low-quality trimming (Martin 2011). Trimmed reads were mapped to the *D. melanogaster* genome r6.15 using *STAR v.2.6* (Dobin *et al.* 2012). On average, 95.9% of the reads mapped to the reference genome. Technical duplications were explored using *dupRadar* (Sayols *et al.* 2016). Overall, we found no bias towards high number of duplicates at low read counts, so we did not remove duplicates from the alignments. We used *featureCounts v.1.6.2* (Liao *et al.* 2014) for counting the number of reads mapping to genes (*reverse-stranded* parameter). Multi-mapping reads and reads overlapping with more than one feature were discarded. RNA-seq coverage values ranged from 153x to 287x. The matrix of counting data was then imported into *DESeq2* (Love *et al.* 2014) for differential expression (DE) analysis following the standard workflow and applying the design formula: *strain + treatment* in the analysis of the tolerant and sensitive strains and only *treatment* in the per-sample DE analysis. Normalisation was performed using the standard DESeq2 normalization method, which accounts for sequencing depth and RNA composition (Anders *et al.* 2010, Love *et al.* 2014).

Differentially expressed genes between treated and control samples were chosen based on both $\log_2$ fold change (≥ 1.0) and adjusted p-values (≤ 0.05). Functional profile analyses of the differentially expressed genes (GO, GSEA and KEGG) were performed using the R package *clusterProfiler* (Yu *et al.* 2012). Breakdown of differentially expressed genes by tissue was performed using the Drosophila Gene Expression Tool (DGET; Hu *et al.* 2017), with gut subsampling data taken from similar aged flies from Marianes and Spradling (2013).

Modulated Modularity Clustering (MMC) was used to group differentially expressed genes into subsets of genetically correlated genes in both treated and control samples. All analysis were carried

out as outlined in Stone *et al*. (2009), except the variance filtering, which was performed in *R v3.5.1* beforehand. Additional network-based analysis was performed using *STRING v10* (Szklarczyk *et al.* 2018) using a minimum interaction score of 0.7. Subsequent visualizations were performed using *Cytoscape v3.7.1* (Lopes *et al.* 2010).

RNAi and Gene Disruption Assays.

Candidate genes were functionally validated using RNAi knockdown lines, Exelixis insertion lines, and one independent deletion mutant (Appendix 1). The Drosophila Genetic Reference Panel (DGRP) lines (Mackay *et al.* 2012), Drosophila Gene Disruption Project (GDP) lines (Bellen 2004) and Transgenic RNAi Project (TRiP) lines (Perkins *et al.* 2015) were all obtained from the Bloomington Drosophila Stock Centre, while additional RNAi lines were obtained from the Vienna Drosophila Resource Centre (VDRC; Dietzl *et al.* 2007).

The choice of *GAL4* driver was based on data obtained for each gene from *FlyAtlas 2* (http://flyatlas.gla.ac.uk/FlyAtlas2/index.html; Leader *et al.* 2017). The *Actin:5C-GAL4* driver was obtained from Bloomington Stock Centre (Stock #4144). The *6g1HR-GAL4* driver, described by Chung *et al*. (2006), and two different background versions of the *MexG-GAL4* driver—originally described by Phillips and Thomas (2006)—were provided by Shane Denecke. As the *Actin:5C-GAL4* is held over the *CyO* balancer the offspring of the crosses who inherited *CyO* were used as controls for their *GAL4* carrying siblings in all instances where this driver was used. As the background of the MexG driver was found to be very week, the knockdown cross with *Jon65Aiv* was repeated with a version backcrossed to *w[1118]*. Kaplan Meier survival analysis was chosen as the best statistical comparison for comparing disrupted and control samples, and all analyses were performed using the R package Survminer (0.4.8). Additionally, relative change in average mortality is also provided as a proxy of the size of the effect of these genes on copper tolerance.

Gut pH Assays

4-5 day old flies from the six lines taken from the RNA-seq analysis were subject to the same assay conditions used for the copper tolerance assays for 24 hours. Assays were performed in triplicate, with each replicate consisting of 30-50 individuals. Higher numbers were required for COR-018 and JUT-008 to account for the level of mortality expected during this timeframe. Flies were then transferred to regular Drosophila media, on which 200ul of a mixture of 1% Bromophenol Blue, dried yeast and water (at 1:1:3 ratio) had been added twenty minutes prior. Flies were permitted to feed for two hours before having their mid-guts dissected in PBS and accessed for loss of acidity. 30-45 samples were dissected from each replicate. Any individuals who proceeded to die after transfer to recovery media were discarded. Samples were determined to have experienced minimal loss in acidity if the cells in the acidic region of the midgut remained yellow (pH<2.3), an intermediate loss if they had faded to

green or brown, and full loss if they could not be distinguished from the surrounding sections (pH >4). No feeding was recorded if no media was present in the gut.
Similar assays were carried out on lines *w[1118]* and the *CG11594* deletion, (*w[1118]; CG11594[1]*) (Appendix 1) over a shorter 16 hour time-period, to account for the greater sensitivity of these lines to copper.

Feeding Avoidance Assays

To measure the effect that the presence of copper has on feeding avoidance, the six lines from the RNA-seq analysis were assayed in similar conditions to that of the tolerance assay, but with the addition of Erioglaucine Disodium Salt (1%, Sigma-Aldrich CAT#861146) to both the treated and control solutions. Eoglaucine Disodium Salt been shown to be an effective tracer up to 48 hours in Drosophila (Shell *et al,* 2018). Assays were performed in triplicate for groups of 25-30 4-7 day old females, with higher numbers used for COR-018 and JUT-008 to account for the degree of mortality expected at the end of this time period. All dead individuals were discarded. Flies were homogenized using a pestle, with each sample consisting of three flies in 620ul of distilled water. After crushing, samples were spun at 14,000rpm for 10 minutes and then frozen for 24 hours. 180ul of supernatant was loaded into each well of a 96 well Nunc-Immuno™ MicroWell™ plate (Sigma-aldrich, CAT#M9410). Measurements were made using a Techan Infinite® 200 Microplate Reader, at 630nm, after 10 seconds of agitation at 9mm. Three technical replicates for six samples, for a total of 18 wells, were loaded for each treatment condition. Each plate contained four water blanks and five standards containing between 0.015 and $1.5E^{-5}$ % of dye. The amount of dye consumed was inferred from a linear model fitted from the points of the standard curve. All results are reported as the fold difference in feeding between treated and control samples for each time point.

Transposable Element Analysis

eQTL analysis. The RNA-Seq data for tolerant and sensitive strains both in control and treated conditions was trimmed using the fastp package (v.0.20.0) (Chen et al. 2018) with default parameters. Expression levels were quantified with the salmon package (v.1.0.0) (Patro et al. 2017) against the ENSEMBL (Dm.BDGP6.22.9) transcripts. Obtained transcripts per million (TPM) were summed up to gene level and rlog normalized using DESeq2 (v.1.28.1) (Love et al. 2014). To test the association between gene expression and TE variants, we used the TE annotations for each one of the six genomes analysed available at https://github.com/sradiouy. The genotype table with the information of the presence/absence of all the TEs present in each one of the strains was created using custom script (https://github.com/sradiouy).
The eQTL analysis was performed using the QTLTools package (v.1.2) (Delaneau et al. 2017). Putative cis-eQTL for the six lines were searched within a 1 kb window around each gene using the cis module in QTLTools. No trans effects were considered. We used the nominal pass to evaluate the

significance of the association of each gene expression level to all the TE insertions within the 1kb window This nominal pass involves the testing of all possible variant-phenotype via linear regression. The variant-phenotype pair with the smallest nominal P-value is kept as the best QTL for that particular TE. In addition, we also performed a permutation pass (100,000 permutations) to adjust for multiple testing. We focused on the significant TE-gene associations with a p-adjusted value < 0.05 and selected the individual gene-TE association with a nominal p-value < 0.05.

Identification of TEs Nearby DEGs. Reference gene annotation was lifted over to each of the six strain assemblies analyzed using Liftoff (v1.4.2), with default parameters, to produce gene annotations of each strain in GFF format. Liftoff annotation was transformed to BED format with a custom python script (https://github.com/sradiouy). Then bedtools closest (v2.29.2) was used to define TE insertions within a 1kb of each gene (-k 10, -D ref) using the TE annotations available at https://github.com/sradiouy.

Phenotypic Validation. TE present and absent outbred populations were constructed for three candidate insertions: *FBti0020195*, *FBti0020057* and *FBti0020036*. Each of these outbred populations was developed to have a mixed genetic background, while remaining consistently homogenous for either the presence or absence of the selected element. For each outbred population, ten females and ten males from each of the five nominated lines (four in the case of *FBti0020195+*) were pooled to establish each population (Appendix 1). Each outbred was maintained for 8 generations in cages before being screened. Tolerance assays were carried out as per the prior experiments, using 10 repeats each for treatment and control. Kaplan Meier survival analysis was performed on present and absence pairs in the same manner as above. Relative change in average mortality is also provided as a proxy of the size of the effect of these genes on copper tolerance

Data availability statement

Data is available in the supplementary tables. Genome assemblies and the raw data (long and short read sequencing) have been deposited in NCBI under the BioProject accession PRJNA55981.RNA-sequence data is available under NCBI accession number: PRJNA646768; GEO: GSE154608.

Acknowledgements

We thank DrosEU members for sharing the European lines, and Shane Denecke and Trent Perry for sharing GAL-4 driver lines. We thank Joshua Schmidt for scripts related to the Kaplan-Meier analysis. We thank Luciano Massetti for making us aware of the availability of the atmospheric pollution data. This project has received funding from the European Research Council (ERC) under

the European Union's Horizon 2020 research and innovation programme (H2020-ERC-2014-CoG-647900).

References

Adriano D. C., 2013 Trace Elements in Terrestrial Environments. Springer Science & Business Media.

Alic N., Andrews T. D., Giannakou M. E., Papatheodorou I., Slack C., Hoddinott M. P., eacute H. M. C., Schuster E. F., Thornton J. M., Partridge L., 2011 Genome-wide dFOXO targets and topology of the transcriptomic response to stress and insulin signalling. Molecular Systems Biology 7: 1–17.

Arthur A. L., Weeks A. R., Sgrò C. M., 2008 Investigating latitudinal clines for life history and stress resistance traits in Drosophila simulansfrom eastern Australia. J Evolution Biol 21: 1470–1479.

Banerjee K. K., Ayyub C., Sengupta S., Kolthur-Seetharam U., 2012 dSir2 deficiency in the fatbody, but not muscles, affects systemic insulin signaling, fat mobilization and starvation survival in flies. Aging (Albany NY) 4: 206–223.

Banerjee K. K., Deshpande R. S., Koppula P., Ayyub C., Kolthur-Seetharam U., 2017 Central metabolic sensing remotely controls nutrient-sensitive endocrine response in Drosophilavia Sir2/Sirt1–upd2–IIS axis. Journal of Experimental Biology 220: 1187–1191.

Barry W. E., Thummel C. S., 2016 The Drosophila HNF4 nuclear receptor promotes glucose-stimulated insulin secretion and mitochondrial function in adults. Elife 5.

Battlay P., Schmidt J. M., Fournier-Level A., Robin C., 2016 Genomic and Transcriptomic Associations Identify a New Insecticide Resistance Phenotype for the Selective Sweep at the Cyp6g1Locus of Drosophila melanogaster. G3 6: 2573–2581.

Bellen H. J., 2004 The BDGP Gene Disruption Project: Single Transposon Insertions Associated With 40% of Drosophila Genes. Genetics 167: 761–781.

Bonilla-Ramirez L., Jimenez-Del-Rio M., Velez-Pardo C., 2011 Acute and chronic metal exposure impairs locomotion activity in Drosophila melanogaster: a model to study Parkinsonism. Biometals 24: 1045–1057.

Calap-Quintana, Pablo; González-Fernández, Javier; Sebastiá-Ortega, Noelia; Llorens, José V.; Moltó, María D. 2017. *Drosophila melanogaster* Models of Metal-Related Human Diseases and Metal Toxicity *Int. J. Mol. Sci.* 18, no. 7: 1456.

Catalán A., Glaser-Schmitt A., Argyridou E., Duchen P., Parsch J., 2016 An Indel Polymorphism in the MtnA 3' Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in Drosophila melanogaster (DJ Begun, Ed.). PLoS Genet 12: e1005987.

Chakraborty M, Emerson JJ, Macdonald SJ, Long AD (2019) Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat Comm* 10, 4872.

Chow C. Y., Wolfner M. F., Clark A. G., 2013 Using natural variation in Drosophila to discover previously unknown endoplasmic reticulum stress genes. Proc Natl Acad Sci USA 110: 9013–9018.

Chung H., Bogwitz M. R., McCart C., Andrianopoulos A., ffrench-Constant R. H., Batterham P., Daborn P. J., 2006 Cis-Regulatory Elements in the Accord Retrotransposon Result in Tissue-Specific Expression of the Drosophila melanogaster Insecticide Resistance Gene Cyp6g1. Genetics 175: 1071–1077.

Cioffi M 1984 Comparative ultrastructure of arthropod transporting epithelia. Amer Zool 24,139–156

Cruz, G.S., Wanderley-Teixeira, V., Antonino, J.D., Gonçalves, G.G.A., Costa, H.N., Ferreira, M.C.N., Neto, C.L., Alves, L.C., Santos, F.A.B. and Teixeira, Á.A.C. 2021, Lufenuron indirectly downregulates *Vitellogenin* in the boll weevil females reducing egg viability. Physiol. Entomol. https://doi.org/10.1111/phen.12337 (early online).

Denecke, S., Fusetto, R., Martelli, F., Giang, A., Battlay, P., Fournier-Level, A., O' Hair, R.A., Batterham, P. (2017). Multiple P450s and Variation in Neuronal Genes Underpins the Response to the Insecticide Imidacloprid in a Population of Drosophila melanogaster. Sci. Rep. 7(1): 11338.

Di Tommaso P., Chatzou M., Floden E. W., Barja P. P., Palumbo E., Notredame C., 2017 Nextflow enables reproducible computational workflows. Nature Publishing Group: 1–4.

Dietzl G., Chen D., Schnorrer F., Su K.-C., Barinova Y., Fellner M., Gasser B., Kinsey K., Oppel S., Scheiblauer S., Couto A., Marra V., Keleman K., Dickson B. J., 2007 A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448: 151–156.

Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T. R., 2012 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Delaneau, O., Ongen H., Brown A. A., Fort A., Panousis N. I., Delaneau O., Dermitzakis E. T., 2017 A complete tool set for molecular QTL discovery and analysis. Nature Communications 8: 1–7.

Egli, D., Yepiskoposyan, H., Selvaraj, A., Balamurugan, K., Rajaram, R., Simons, A., Multhaup, G., Mettler, S., Vardanyan, A., Georgiev, O., Schaffner, W. 2006. A family knockout of all four Drosophila metallothioneins reveals a central role in copper homeostasis and detoxification. Mol. Cell. Biol. 26(6): 2286--2296.

Everman E. R., Cloud-Richardson K. M., Macdonald S. J., Characterizing the genetic basis of copper toxicity in *Drosophila* reveals a complex pattern of allelic, regulatory, and behavioral variation Genetics, 2021; iyaa020, https://doi.org/10.1093/genetics/iyaa020

Ewels P. A., Peltzer A., Fillinger S., Patel H., Alneberg J., Wilm A., Garcia M. U., Tommaso P., Nahnsen S., 2020 The nf-core framework for community-curated bioinformatics pipelines. Nature Biotechnology: 1–3.

Fabian D. K., Kapun M., Nolte V., Kofler R., Schmidt P. S., Schlötterer C., Flatt T., 2012 Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogasterfrom North America. Molecular Ecology 21: 4748–4769.

Fabregat A., Jupe S., Matthews L., Sidiropoulos K., Gillespie M., Garapati P., Haw R., Jassal B., Korninger F., May B., Milacic M., Roca C. D., Rothfels K., Sevilla C., Shamovsky V., Shorser S., Varusai T., Viteri G., Weiser J., Wu G., Stein L., Hermjakob H., D'Eustachio P., 2017 The Reactome Pathway Knowledgebase. Nucleic Acids Research 46: D649–D655.

Fernandez-Falcon M., Borges-Perez A., Perez-Frances J. F., Borges A. A., Lopez-Carreno I., 1994 Available micronutrients in agricultural soils of Tenerife (Canary Islands). I.: copper and zinc. Agrochima: 268–276.

Filshie B. K., Poulson D. F., Waterhouse D. F., 1971 Ultrastructure of the copper-accumulating region of the Drosophila larval midgut. Tissue Cell 3: 77–102.

Green L., Battlay P., Fournier-Level A., Good R. T., Robin C., 2019 Cis- and trans-acting variants contribute to survivorship in a naïve Drosophila melanogasterpopulation exposed to ryanoid insecticides. Proc Natl Acad Sci USA 63: 201821713.

Guio L., Barrón M. G., González J., 2014 The transposable element Bari-Jhehmediates oxidative stress response in Drosophila. Molecular Ecology 23: 2020–2030.

Hallas R., Schiffer M., Hoffmann A. A., 2002 Clinal variation in Drosophila serrata for stress resistance and body size. Genet Res Apr;79(2):141-8

Hoffmann A. A., Weeks A. R., 2006 Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica 129: 133–147.

Horváth V., Merenciano M., González J., 2017 Revisiting the Relationship between Transposable Elements and the Eukaryotic Stress Response. Trends in Genetics 33: 832–841.

Hu Y., Comjean A., Perrimon N., Mohr S. E., 2017 The Drosophila Gene Expression Tool (DGET) for expression analyses. BMC Bioinformatics 18, 98:

Hung R.-J., Hu Y., Kirchner R., Liu Y., Xu C., Comjean A., Tattikota S. G., Li F., Song W., Ho Sui S., Perrimon N., 2020 A cell atlas of the adult Drosophila midgut. Proc Natl Acad Sci USA 117: 1514–1523.

Johnson R. M., Dahlgren L., Siegfried B. D., Ellis M. D., 2013 Acaricide, Fungicide and Drug Interactions in Honey Bees (Apis mellifera) (NE Raine, Ed.). PLoS ONE 8: e54092.

Kalsi M., Palli S. R., 2017 Transcription factor cap n collar C regulates multiple cytochrome P450 genes conferring adaptation to potato plant allelochemicals and resistance to imidacloprid in Leptinotarsa decemlineata (Say). Insect Biochemistry and Molecular Biology **83**: 1–12.

Kapun M, Barron Aduriz MG, Staubach F, Vieira J, Obbard D, Goubert C, et al. 2020 Genomic analysis of European Drosophila populations reveals longitudinal structure and continent-wide selection. *Mol. Biol. Evol.*, 37: 2661–2678

King-Jones K., Horner M. A., Lam G., Thummel C. S., 2006 The DHR96 nuclear receptor regulates xenobiotic responses in Drosophila. Cell Metabolism 4: 37–48.

Kuraishi T., Binggeli O., Opota O., Buchon N., Lemaitre B., 2011 Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. Proc Natl Acad Sci USA 108: 15966–15971.

Leader D. P., Krause S. A., Pandit A., Davies S. A., Dow J. A. T., 2017 FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. Nucleic Acids Research 46: D809–D815.

Le Manh, H., Guio, L., Merenciano, M., Rovira, Q., Barrón, M.G., González, J. 2017. Natural and laboratory mutations in kuzbanian are associated with zinc stress phenotypes in Drosophila melanogaster. Sci. Rep. 7: 42663.

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

Lopes C. T., Franz M., Kazi F., Donaldson S. L., Morris Q., Bader G. D., 2010 Cytoscape Web: an interactive web-based network browser. Bioinformatics 26: 2347–2348.

Love M. I., Huber W., Anders S., 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 31.

Mackay T. F. C., Richards S., Stone E. A., Barbadilla A., Ayroles J. F., Zhu D., Casillas S., Han Y., Magwire M. M., Cridland J. M., Richardson M. F., Anholt R. R. H., Barrón M., Bess C., Blankenburg K. P., Carbone M. A., Castellano D., Chaboub L., Duncan L., Harris Z., Javaid M., Jayaseelan J. C., Jhangiani S. N., Jordan K. W., Lara F., Lawrence F., Lee S. L., Librado P., Linheiro R. S., Lyman R. F., Mackey A. J., Munidasa M., Muzny D. M., Nazareth L., Newsham I., Perales L., Pu L.-L., Qu C., Ràmia M., Reid J. G., Rollmann S. M., Rozas J., Saada N., Turlapati L., Worley K. C., Wu Y.-Q., Yamamoto A., Zhu Y., Bergman C. M., Thornton K. R., Mittelman D., Gibbs R. A., 2012 The Drosophila melanogaster Genetic Reference Panel. Nature 482: 173–178.

Marianes A., Spradling A. C., 2013 Physiological and stem cell compartmentalization within the Drosophila midgut. Elife 2: 155.

Maroni G., Wise J., Young J. E., Otto E., 1987 Metallothionein gene duplications and metal tolerance in natural populations of Drosophila melanogaster. Genetics 117: 739–744.

Martin M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12.

Massadeh A., Al-Momani F., Elbetieha A., 2008 Assessment of Heavy Metals Concentrations in Soil Samples from the Vicinity of Busy Roads: Influence on Drosophila melanogaster Life Cycle. Biol Trace Elem Res 122: 292–299.

Mateo, L., Ullastres, A., González, J. 2014 A transposable element insertion confers xenobiotic resistance in Drosophila. PLoS Genet. 10(8): e1004560.

Mateo, L., Rech, G.E. & González, J. 2018 Genome-wide patterns of local adaptation in Western European *Drosophila melanogaster* natural populations. *Sci Rep* **8,** 16143. https://doi.org/10.1038/s41598-018-34267-0

McNulty M., Puljung M., Jefford G., Dubreuil R., 2001 Evidence that a copper-metallothionein complex is responsible for fluorescence in acid-secreting cells of the Drosophila stomach. Cell and Tissue Research 304: 383–389.

Merenciano, M., Ullastres, A., de Cara, M.A., Barrón, M.G., González, J. 2016. Multiple Independent Retroelement Insertions in the Promoter of a Stress Response Gene Have Variable Molecular and Functional Effects in Drosophila. PLoS Genet. 12(8): e1006249.

Mercer J. F. B., Camakaris J., 2004 *Copper: the Jekyll and Hyde Element* . Australian Biochemist 35: 5–8.

Merritt TJS and Bewick AJ 2017. Genetic Diversity in Insect Metal Tolerance. *Front. Genet.* 8:172.

Navarro JA, Schneuwly S. 2017 *Copper and Zinc Homeostasis: Lessons from Drosophila melanogaster*. Front Genet. Dec 21;8:223.

Norgate M., Adam Southon, Zou S., Zhan M., Sun Y., Batterham P., camakaris J., 2007 Copper homeostasis gene discovery in Drosophila melanogaster. Biometals 20: 683–697.

Norgate M., Lee E., Southan, Farlow A., Batterham P., Camakaris J., Burke R., 2006 Essential roles in development and pigmentation for the Drosophila copper transporter DmATP7. Mol Biol Cell 17: 475–484.

Ojima N., Hara Y., Ito H., Yamamoto D., 2018 Genetic dissection of stress-induced reproductive arrest in Drosophila melanogaster females (PH Taghert, Ed.). PLoS Genet 14: e1007434.

Orgiazzi A., Ballabio C., Panagos P., Jones A., Fernández-Ugalde O., 2017 LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur J Soil Sci 69: 140–153.

Palu R. A. S., Thummel C. S., 2016 Sir2 Acts through Hepatocyte Nuclear Factor 4 to maintain insulin Signaling and Metabolic Homeostasis in Drosophila (SK Kim, Ed.). PLoS Genet 12: e1005978.

Panagos P., Ballabio C., Lugato E., Jones A., Borrelli P., Scarpa S., Orgiazzi A., Montanarella L., 2018 Potential Sources of Anthropogenic Copper Inputs to European Agricultural Soils. Sustainability 10: 2380.

Perkins L. A., Holderbaum L., Tao R., Hu Y., Sopko R., McCall K., Yang-Zhou D., Flockhart I., Binari R., Shim H.-S., Miller A., Housden A., Foos M., Randkelv S., Kelley C., Namgyal P., Villalta C., Liu L.-P., Jiang X., Huan-Huan Q., Wang X., Fujiyama A., Toyoda A., Ayers K., Blum A., Czech B., Neumuller R., Yan D., Cavallaro A., Hibbard K., Hall D., Cooley L., Hannon G. J., Lehmann R., Parks A., Mohr S. E., Ueda R., Kondo S., Ni J.-Q., Perrimon N., 2015 The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. Genetics 201: 843–852.

Phillips M. D., Thomas G. H., 2006 Brush border spectrin is required for early endosome recycling in Drosophila. Journal of Cell Science 119: 1361–1370.

Pölkki M., Rantala M. J., 2021 Exposure to copper during larval development has intra- and trans-generational infuence on fitness in later life . Ecotoxicology and Environmental Safety 207: 111133.

Quinlan A. R., Hall I. M., 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Rech G. E., Bogaerts-Márquez M., Barrón M. G., Merenciano M., Villanueva-Cañas J. L., Horváth V., Fiston-Lavier A.-S., Luyten I., Venkataram S., Quesneville H., Petrov D. A., González J., 2019 Stress response, behavior, and development are shaped by transposable element-induced mutations in Drosophila (S Schaack, Ed.). PLoS Genet 15: e1007900.

Romic M., Romic D., 2003 Heavy metals distribution in agricultural topsoils in urban area. Env Geol 43: 795–805.

Ross, J., Jiang, H., Kanost, M.R., Wang, Y. 2003. Serine proteases and their homologs in the Drosophila melanogaster genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene 304(1-2): 117--131.

Roelofs D, Janssens TK, Timmermans MJ, Nota B, Mariën J, Bochdanovits Z, Ylstra B, Van Straalen NM. 2009. Adaptive differences in gene expression associated with heavy metal tolerance in the soil arthropod Orchesella cincta. Mol Ecol. Aug;18(15):3227-39.

Sayols S., Scherzinger D., Klein H., 2016 dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. BMC Bioinformatics: 1–5.

Schmidt J. M., Robin C., 2011a An Adaptive Allelic Series Featuring Complex Gene Rearrangements (DJ Begun, Ed.). PLoS Genet 7: e1002347.

Schmidt J. M., Robin C., 2011b An Adaptive Allelic Series Featuring Complex Gene Rearrangements (DJ Begun, Ed.). PLoS Genet 7: e1002347.

Schmidt P. J., Kunst C., Culotta V. C., 2000 Copper Activation of Superoxide Dismutase 1 (SOD1) in Vivo. J. Biol. Chem. 275: 33771–33776.

Shell, B.C., Schmitt, R.E., Lee, K.M. *et al.* 2018 Measurement of solid food intake in *Drosophila* via consumption-excretion of a dye tracer. *Sci Rep* **8,** 11536

Song M. O., Freedman J. H., 2011 Role of hepatocyte nuclear factor 4α in controlling copper-responsive transcription. Biochim. Biophys. Acta 1813: 102–108.

Southon A., Burke R., Norgate M., Batterham P., Camakaris J., 2004 Copper homoeostasis in Drosophila melanogaster S2 cells. Biochem. J. 383: 303–309.

Southon A., Farlow A., Norgate M., Burke R., Camakaris J., 2008 Malvolio is a copper transporter in Drosophila melanogaster. Journal of Experimental Biology 211: 709–716.

Stone E. A., Ayroles J. F., 2009 Modulated modularity clustering as an exploratory tool for functional genomic inference. PLoS Genet 5: e1000479.

Sun Y., Liu C., Liu Y., Hosokawa T., Saito T., Kurasaki M., 2014 Changes in the expression of epigenetic factors during copper-induced apoptosis in PC12 cells. Journal of Environmental Science and Health, Part A 49: 1023–1028.

Szklarczyk D., Gable A. L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N. T., Morris J. H., Bork P., Jensen L. J., Mering C. V., 2018 STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research 47: D607–D613.

Tapp R. L., Hockaday A., 1977 Combined histochemical and x-ray microanalytical studies on the copper-accumulating granules in the mid-gut of larval Drosophila. Journal of Cell Science 26: 201–215.

Terashima J., Bownes M., 2005 A microarray analysis of genes involved in relating egg production to nutritional intake in Drosophila melanogaster. Cell Death Differ 12: 429–440.

Thomas P., Kejariwal A., Guo N., Mi H., Campbell M. J., Muruganujan A., and Lazareva-Ulitsky, B. 2006. Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res. Jun: 1;34

Thomas J., Vadnagara K., Pritham E. J., 2014 DINE-1, the highest copy number repeats in Drosophila melanogaster are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). Mob DNA 5: 18.

Turski M. L., Thiele D. J., 2007 Drosophila Ctr1A functions as a copper transporter essential for development. J. Biol. Chem. 282: 24017–24026.

Ullastres A., Merenciano M.; Gonzalez J. 2019 Natural transposable element insertions drive expression changes in genes underlying Drosophila immune response. *Biorxiv.*

Wat LW, Chao C, Bartlett R, Buchanan JL, Millington JW, et al. 2020 A role for triglyceride lipase *brummer* in the regulation of sex differences in *Drosophila* fat storage and breakdown. PLOS Biology 18(1): e3000595. https://doi.org/10.1371/journal.pbio.3000595

Weber A. L., Khan G. F., Magwire M. M., Tabor C. L., Mackay T. F. C., Anholt R. R. H., 2012 Genome-Wide Association Analysis of Oxidative Stress Resistance in Drosophila melanogaster (T Flatt, Ed.). PLoS ONE 7: e34745.

Wilson T. G., 2005 Drosophila: sentinels of environmental toxicants. Integr. Comp. Biol. 45: 127–136.

Yepiskoposyan H., Egli D., Fergestad T., Selvaraj A., Treiber C., Multhaup G., Georgiev O., Schaffner W., 2006 Transcriptome response to heavy metal stress in Drosophila reveals a new zinc transporter that confers resistance to zinc. Nucleic Acids Research 34: 4866–4877.

Yu G., Wang L.-G., Han Y., He Q.-Y., 2012 clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology 16: 284–287.

Zamberlan, D.C., Halmenschelager, P.T., Silva, Lui, Rocha, Joao Batista. 2019. Copper decreases associative learning and memory in Drosophila melanogaster. Science of The Total Environment. 710. 135306.

Table 1. Differentially Expressed Genes up-regulated in the tolerant and down-regulated in the sensitive strains. ND is no data.

Gene name	Biological process	Molecular function	Prior copper response association
CG13078	ND	heme binding, oxidoreductase activity, transmembrane ascorbate ferrireductase activity	Within QTL (Everman et al. 2021)
CG31091	lipid metabolic process	sterol esterase activity, triglyceride lipase activity	ND
Hml	hemolymph coagulation, hemostasis, wound healing	protein homodimerization activity, chitin binding	Up-regulated in control vs treated (Everman et al. 2020)
Jon99Ci	proteolysis	endopeptidase activity, serine-type endopeptidase activity	ND
Lectin 24Db	ND	fucose binding, glycosylated region protein binding, mannose binding	Downregulated in control vs treated (Everman et al. 2020)
NtR	chemical synaptic transmission, ion transmembrane transport, nervous system process, regulation of membrane potential, signal transduction	extracellular ligand-gated ion channel activity, neurotransmitter receptor activity, transmembrane signaling receptor activity	ND