THE COMPLEX GENETICS OF LIFESPAN AND XENOBIOTIC RESISTANCE IN *DROSOPHILA MELANOGASTER*

By

Chad Allen Highfill

Submitted to the graduate degree program in Molecular Biosciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson – Stuart Macdonald

_____
Brian Ackley

_____
Eric Deeds

_____
Erik Lundquist

_____
Robert Ward

_____
John Kelly

Date Defended: July 5, 2016

The Dissertation Committee for Chad Allen Highfill
certifies that this is the approved version of the following dissertation:

THE COMPLEX GENETICS OF LIFESPAN AND XENOBIOTIC RESISTANCE IN *DROSOPHILA MELANOGASTER*

_____
Chairperson – Stuart Macdonald

Date approved: July 5, 2016

Abstract

It is constantly observed that populations of species, like the DSPR, harbor considerable genetic variation for phenotypic traits. However, our understanding of the location, effect, and frequency of alleles that create variable genetic effects for each phenotypic trait is poorly understood and remains elusive. To tease apart the location, effect, and frequency of alleles with variable effects on phenotypic traits, this dissertation employs the DSPR, comprised of the elite model system *Drosophila melanogaster*, to investigate lifespan and xenobiotic resistance. Experiments on lifespan uncovered multiple quantitative trait loci (QTL) that harbor genes that shape lifespan and small sets of expression candidates. Moreover, as many expression studies have discovered, our expression candidates are enriched with antimicrobial defense genes that increase in expression with age, while electron transport chain genes decrease in expression with age. Exploring xenobiotic resistance we identify that *Cyp28d1* and *Cyp28d2* are likely to have variable effects on nicotine resistance, and *Ugt86Dd* is functionally important for nicotine resistance. Lastly, we discovered a complex 22bp deletion in *Ugt86Dd* that our data suggest is likely a causative variant within *Ugt86Dd* contributing variable effects on nicotine resistance in the DSPR. Our studies demonstrate the genetic architecture of lifespan is more complex than what is reported and lose of function variants may have an important role in creating variable effects on xenobiotic resistance.

Acknowledgements

To recognize important moments and people that have been inspirational toward my success. I would like to think my parents, Norma J. and Jay D. Highfill for encouraging me to pursue my dreams. Even though we didn't always have a lot of money for the best things, I managed to thrive. To my dad, thank you for these words "Chad, you are a pretty fart smeller, and I believe you can be whatever you set your mind to be." Since I was a freshman in high school, that simple sentence has shaped my life in profound ways. I also want to personal thank, Kimiko and Shuji, for helping reduce my family's financial pressure. This greatly helped me finish my PhD.

I want to give a special thank you to Mrs. Kohler and Mrs. McPheeters for always sparking my interest in science and in myself. Mrs. Kohler, without your guidance as a small child, I wouldn't have learned how to set goals. I hated being in a special education class room because I was picked on for being different, and all I wanted was to be included with everyone else. You told me "if I worked really hard, I could make it into the mainstream classroom." I did that by learning to harness my ADHD in a positive way through science experiments, of which my first was our plant experiment. This is where my scientific spark was created! Now to Mrs. McPheeters, your exploratory science class was the final catalyst that sent me down this amazing path. You two have had the most impact in my life toward achieving this goal.

To, the JTS, Heather Shinogle, AJ Merlott, and Shara Thati for the great times we had throughout our graduate education. You guys have always been there for me and helped me generate ideas unknowing. I also want to thank Christian Ray for being an awesome scientific friend and colleague. You have shown me what it takes to think outside to box and fight the good science fight. To John Kelly, for teaching the best college course I have taken from

undergraduate to graduate school. You have an amazing way of taking the most complex material and making it very simple to understand and in the process have shown me how important statistics are.

To Stuart J. Macdonald, for teaching me everything I know about quantitative genetics and about doing large experiments, statistically analysis, programming, scientific writing, and professionalism. I deeply respect you as a scientist and I hope to see you around in the community.

To my beautiful wife, Mana Uchida, you are my best friend and love of my life. I enjoy every moment with you and thank you for giving me the time and financial means to chase after my dreams. I know it has been rough, but we are at an end to this chapter. It is time to start a new chapter were anything is possible.

To the most important person in my life, my son Kai Uchida Highfill, from the first second I got to hold you, I realized my life will never be the same in the best way possible. I have gotten to see you grow up through you first year of life. Let me tell you, I have seen an amazing young man that is full of great potential. You have demonstrated time and time that you are incredibly strong and smart. I believe you will be able to accomplish anything you set your mind to. Please remember the quotes I mentioned earlier and may they serve you well as they did me. Regardless if I am around or not around, always question the world around you and only accept the best from people. Lastly, if it wasn't for you, I couldn't have finished my PhD. I hope son, that I can leave the world a better place for you. Remember, "You can always achieve what you set your mind too and never give up."

Table of Contents

Page Numbers

List of Figures

Figures **Page Numbers**

List of Tables

Chapter I

Introduction to Natural Variation and Projects

Introduction

In nature, evolutionary processes such as genetic drift, gene flow, founder effects, bottlenecks, mutations, non-random mating, and random mating create natural genetic variation within populations, between populations, and species. Currently, this natural genetic variation affecting an array of traits is poorly understood at the molecular level due to low heritability's, lack of ability to phenotype traits, and the complexity of the genetic architecture. With these pitfalls there are a plethora of traits that can be mapped to the molecular level. However, the traditional mapping methodologies, association mapping and linkage mapping, have had limited success in mapping variation to the molecular level leaving many polymorphisms conferring an affect elusive [1–8].

Linkage mapping was first noted over 90 years ago [9], this approach has strong statistical power to detect genetic variation, but suffers from limited resolution [10] and is limited in the amount of segregating variation that exist between the initial founders. In comparison, association mapping allows for great mapping resolution, but requires large numbers of individuals to uncover genetic variation with effects of <5% to the overall phenotype [11], suffers from multiple test correction problems, and reproducibility [12]. To maintain the strengths and limit the weakness of each of these two approaches, advanced intercross lines can be developed [13]. In this ideology, many founders are encouraged mate randomly over several generations to propagate the plethora of genetic diversity and the number of recombinational events to reduce linkage disequilibrium throughout the genome. Recently, this ideology was employed to construct several multi-parental mapping panels that include the *Arabidopsis thaliana* MAGIC (Multiparent Advanced Generation Inter-Cross) lines [14], the maize NAM (Nested Association Mapping population) panel [15], the mouse collaborative cross [16], the

Diversity Outbred mouse population [17], and the DSPR (*Drosophila* Synthetic Population Resource) [18,19]. The DSPR is a resource composed of ~1600 RILs (recombinant inbred lines) that are fine mosaics of the original 15 founders sampled worldwide. The composition of the DSPR allows QTL (quantitative trait loci) to not only be mapped to narrow genomic intervals in comparison to traditional approaches, but also enables small effect QTL to be mapped [18] and the ability to estimate the effect frequency of mapped QTL. Moreover, these approaches, including the DSPR, relies on high-throughput sequencing of the founders and mapping panel. Thus, only a time commitment in phenotyping is required for any trait of interest. We did exactly this by phenotyping 805 RILs over a course of five blocks for longevity.

Variation in longevity is a classic life history trait that has a direct impact on an organism's life table in nature. Of interest is the amount of variation is attributed to genetic and/or environmental factors. Several studies have used various mapping designs and attempted to elucidate an understanding to how much of this variation is attributed to genetic and/or environmental factors [2,5,6,20–28]; however, we are the first study to investigate variation in lifespan with a mapping panel that was created from multiple parents and employing 805 RILs. With the DSPR ability to map QTL of small effect to narrow genetic intervals, we attempted to characterize the complex genetic architecture of longevity in mated females of the DSPR. We mapped 5 QTL that contained a modest number of genes, which collectively explained 22.2% of the genetic variation for lifespan in the DSPR. Furthermore, the mapped genomic intervals recapitulated previously implicated genomic regions identified in various lifespan QTL screens. Follow-up RNAseq studies in heads and bodies confirmed several known and novel candidate lifespan genes under our mapped QTL, and a GO (Gene Ontology) analysis discovered an enrichment genes that function in category's such as defense response, egg coat formation, myofibril assembly, and electron transport chain. Several studies have used array-based

expression profiling to identify genes that change with age in various *D. melanogaster* populations, and we sought to determine the extent of overlap in the genes identified among studies [27,29–32]. Overall, 83% of the genes we identify as differentially expressed in bodies were identified in at least one other study, and 59% of the genes we identify in heads replicated. A total of 55 RNAseq candidates are present within QTL. Our results suggest an exceedingly complex genetic architecture for lifespan in *Drosophila*.

While several studies have successfully identified genomic intervals important for various drug resistance traits via QTL mapping in various organisms [33–39], resolving these genomic intervals to the respective causative gene(s) is challenging. This dissertation is the first to employ an array of functional assays to resolve nicotine resistance QTL, specifically QTL1 and QTL4, previously mapped by Marriage *et al*. 2014 to the likely causative genes. We utilized QCTs (quantitative complementation tests) and RNAi to demonstrate *Cyp28d1*, *Cyp28d2*, and *Ugt86Dd* have functional roles in nicotine resistance. During the course of this study a complex 22bp InDel was discovered in the second exon of *Ugt86Dd* that causes a frameshift mutation that results in a premature stop. Moreover, over-expression and CRISPR assays provide direct evidence that *Ugt86Dd* is functionally important and provide indirect evidence that this InDel is likely causative. We provide a framework that allows one to resolve the genomic interval to the causative gene(s) and will increase confidence for investigators to elucidate the functional variants under mapped QTL.

Literature Cited

1. Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH, Mackay TF. High resolution mapping of genetic factors affecting abdominal bristle number in Drosophila melanogaster. Genetics. 1995;139:1273–91.

2. Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF. Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 1997;94:9734–9.

3. Gurganus MC, Fry JD, Nuzhdin SV, Pasyukova EG, Lyman RF, Mackay TF. Genotype-environment interaction at quantitative trait loci affecting sensory bristle number in Drosophila melanogaster. Genetics. 1998;149:1883–98.

4. Gurganus MC, Nuzhdin SV, Leips JW, Mackay TF. High-resolution mapping of quantitative trait loci for sternopleural bristle number in Drosophila melanogaster. Genetics. 1999;152:1585–604.

5. Leips J, Mackay TFC. Quantitative Trait Loci for Life Span in Drosophila melanogaster: Interactions With Genetic Background and Larval Density. Genetics. 2000;155:1773–88.

6. Vieira C, Pasyukova EG, Zeng Z-B, Hackett JB, Lyman RF, Mackay TFC. Genotype-environment interaction for quantitative trait loci affecting life span in Drosophila melanogaster. Genetics. 2000;154:213–27.

7. Unckless RL, Rottschaefer SM, Lazzaro BP. A Genome-Wide Association Study for Nutritional Indices in Drosophila. G3 GenesGenomesGenetics. 2015;5:417–25.

8. Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. Annu. Rev. Pathol. 2011;6:193–222.

9. Sax K. The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS. Genetics. 1923;8:552–60.

10. Mackay TFC. Quantitative Trait Loci In Drosophila. Nat. Rev. Genet. 2001;2:11–20.

11. Long AD, Langley CH. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. 1999;9:720–31.

12. NCI-NHGRI Working Group on Replication in Association Studies, Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, et al. Replicating genotype-phenotype associations. Nature. 2007;447:655–60.

13. Darvasi A, Soller M. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics. 1995;141:1199–207.

14. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, et al. A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 2009;5:e1000551.

15. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. Science. 2009;325:714–8.

16. Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS, et al. Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Res. 2011;21:1213–22.

17. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 2004;36:1133–7.

18. King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, et al. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 2012;22:1558–66.

19. King EG, Macdonald SJ, Long AD. Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics. 2012;191:935–49.

20. Curtsinger JW, Khazaeli AA. Lifespan, QTLs, age-specificity, and pleiotropy in Drosophila. Mech. Ageing Dev. 2002;123:81–93.

21. Reiwitch SG, Nuzhdin SV. Quantitative trait loci for lifespan of mated Drosophila melanogaster affect both sexes. Genet Res. 2002;80:225–30.

22. Geiger-Thornsberry GL, Mackay TFC. Quantitative trait loci affecting natural variation in Drosophila longevity. Mech. Ageing Dev. 2004;125:179–89.

23. Forbes SN, Valenzuela RK, Keim P, Service PM. Quantitative trait loci affecting life span in replicated populations of Drosophila melanogaster. I. Composite interval mapping. Genetics. 2004;168:301–11.

24. Nuzhdin SV, Khazaeli AA, Curtsinger JW. Survival Analysis of Life Span Quantitative Trait Loci in Drosophila melanogaster. Genetics. 2005;170:719–31.

25. Wang M-H, Lazebny O, Harshman LG, Nuzhdin SV. Environment-dependent survival of Drosophila melanogaster: a quantitative genetic analysis. Aging Cell. 2004;3:133–40.

26. Wilson RH, Morgan TJ, Mackay TFC. High-Resolution Mapping of Quantitative Trait Loci Affecting Increased Life Span in Drosophila melanogaster. Genetics. 2006;173:1455–63.

27. Lai C-Q, Parnell LD, Lyman RF, Ordovas JM, Mackay TFC. Candidate genes affecting Drosophila life span identified by integrating microarray gene expression analysis and QTL mapping. Mech. Ageing Dev. 2007;128:237–49.

28. Defays R, Gomez FH, Sambucetti P, Scannapieco AC, Loeschcke V, Norry FM. Quantitative trait loci for longevity in heat-stressed Drosophila melanogaster. Exp. Gerontol. 2011;46:819–26.

29. Pletcher SD, Macdonald SJ, Marguerie R, Certa U, Stearns SC, Goldstein DB, et al. Genome-wide transcript profiles in aging and calorically restricted Drosophila melanogaster. Curr Biol. 2002;12:712–23.

30. Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, et al. Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 2004;101:7663–8.

31. Zhan M, Yamaza H, Sun Y, Sinclair J, Li H, Zou S. Temporal and spatial transcriptional profiles of aging in Drosophila melanogaster. Genome Res. 2007;17:1236–43.

32. Carlson KA, Gardner K, Pashaj A, Carlson DJ, Yu F, Eudy JD, et al. Genome-Wide Gene Expression in relation to Age in Large Laboratory Cohorts of Drosophila melanogaster. Genet. Res. Int. 2015;2015:835624.

33. Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FH. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan Anopheles gambiae associated with resistance to DDT and pyrethroids. Insect Mol. Biol. 2000;9:491–7.

34. Defays R, Bertoli CI. Quantitative trait loci for response to ethanol in an intercontinental set of recombinant inbred lines of Drosophila melanogaster. Alcohol Fayettev. N. 2012;46:737–45.

35. Witzig TE, Laumann KM, Lacy MQ, Hayman SR, Dispenzieri A, Kumar S, et al. A phase III randomized trial of thalidomide plus zoledronic acid versus zoledronic acid alone in patients with asymptomatic multiple myeloma. Leukemia. 2013;27:220–5.

36. King EG, Kislukhin G, Walters KN, Long AD. Using Drosophila melanogaster To Identify Chemotherapy Toxicity Genes. Genetics. 2014;198:31–43.

37. Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. Genetics. 2014;198:45–57.

38. Reyes-Solis G del C, Saavedra-Rodriguez K, Suarez AF, Iv WCB. QTL Mapping of Genome Regions Controlling Temephos Resistance in Larvae of the Mosquito Aedes aegypti. PLOS Negl Trop Dis. 2014;8:e3177.

39. Najarro MA, Hackett JL, Smith BR, Highfill CA, King EG, Long AD, et al. Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. PLoS Genet. 2015;11:e1005663.

Chapter II

Genetic analysis of variation in lifespan using a multiparental advanced intercross *Drosophila* mapping population

ABSTRACT

Background

Considerable natural variation for lifespan exists within human and animal populations. Genetically dissecting this variation can elucidate the pathways and genes involved in aging, and help uncover the genetic mechanisms underlying risk for age-related diseases. Studying aging in model systems is attractive due to their relatively short lifespan, and the ability to carry out programmed crosses under environmentally-controlled conditions. Here we investigated the genetic architecture of lifespan using the *Drosophila* Synthetic Population Resource (DSPR), a multiparental advanced intercross mapping population.

Results

We measured lifespan in females from 805 DSPR lines, mapping five QTL (Quantitative Trait Loci) that each contribute 4-5% to among-line lifespan variation in the DSPR. Each of these QTL co-localizes with the position of at least one QTL mapped in 13 previous studies of lifespan variation in flies. However, given that these studies implicate >90% of the genome in the control of lifespan, this level of overlap is unsurprising. DSPR QTL intervals harbor 11-155 protein-coding genes, and we used RNAseq on samples of young and old flies to help resolve pathways affecting lifespan, and identify potentially causative loci present within mapped QTL intervals. Broad age-related patterns of expression revealed by these data recapitulate results from previous work. For example, we see an increase in antimicrobial defense gene expression with age, and a decrease in expression of genes involved in the electron transport chain. Several genes within

QTL intervals are highlighted by our RNAseq data, such as *Relish*, a critical immune response gene, that shows increased expression with age, and *UQCR-14*, a gene involved in mitochondrial electron transport, that has reduced expression in older flies.

Conclusions

The five QTL we isolate collectively explain a considerable fraction of the genetic variation for female lifespan in the DSPR, and implicate modest numbers of genes. In several cases the candidate loci we highlight reside in biological pathways already implicated in the control of lifespan variation. Thus, our results provide further evidence that functional genetics tests targeting these genes will be fruitful, lead to the identification of natural sequence variants contributing to lifespan variation, and help uncover the mechanisms of aging.

BACKGROUND

Life expectancy in developed countries has markedly increased in the last 100 years, and individuals born in the USA in 2011 can expect to live to nearly 80 years old [1]. Since old age is a major risk factor for an array of diseases [2], the prevalence of age-related disorders is concomitantly increasing as populations age. Given the significant segregating genetic variation for lifespan within populations [3], with twin studies indicating modest heritabilities of approximately 20-30% [4, 5], a key challenge for biomedical science is to understand the genetic basis of variation in lifespan, and articulate any mechanistic relationships between aging and the risk for age-related disease.

To localize genes and/or variants associated with age in humans researchers have frequently used a GWAS (Genomewide Association Study) approach, comparing a cohort of centenarians to a cohort of middle-aged controls. Studies of this type have repeatedly associated age with variation at the *APOE* locus [6-8], a gene also known to strongly influence risk for Alzheimer's [9]. However, such studies are often small due to the difficulty obtaining large cohorts of aged individuals, and thus lack power [10]. They also encounter the same problems as all GWAS, in that rare causative variants, and genes that segregate for a heterogeneous set of disease-causing alleles, are essentially invisible to the standard analytical methods employed [11-13]. In addition, direct genetic analysis of aging in humans must be carried out in the face of considerable environmental heterogeneity among samples.

One alternative fruitful strategy to discover the genetic and environmental determinants of variation in aging has been to use model systems, where total lifespan is much shorter than in

humans, powerful genetic mapping experiments can be carried out using specifically bred individuals, *in vivo* genetic manipulation is possible, the environment throughout lifespan can be regulated to a large degree, and environmental interventions can be evaluated easily. Work in a number of non-human systems - from yeast, to flies, to mice - has demonstrated that dietary restriction routinely extends lifespan [14], and trials of dietary restriction in humans have yielded beneficial health responses [15, 16]. In addition, mutations in members of the insulin signaling pathway show robust effects on lifespan in several systems, such as *C. elegans* [17, 18], *Drosophila* [19], and mice [20]. Such observations suggest shared physiological mechanisms may underlie the response to aging, and imply some level of conservation in the genetic mechanisms contributing to lifespan variation.

In model systems, two broad strategies can be implemented to identify genes and pathways impacting lifespan and age-related phenotypes: Mutational analyses, and mapping loci contributing to variation in lifespan in natural, or semi-natural laboratory populations. Given the relative ease with which large-effect mutations can be generated and interrogated in flies, multiple studies have screened large sets of induced mutations for their effects on lifespan (e.g., [21, 22]), and detailed mechanistic studies targeting specific genes and pathways have added considerably to our understanding of the aging process. However, such loci may be distinct from those that harbor naturally-segregating sites underlying variation in lifespan (compare tables 1, 2, and 3 in [23]). To identify genes contributing to natural variation in lifespan, *Drosophila* researchers have used techniques such as QTL (Quantitative Trait Locus) mapping [24] to screen the genome in an unbiased fashion, and - coupled with downstream functional tests - have successfully implicated a small number of genes in the control of lifespan variation (e.g., *Dopa decarboxylase*, [25]).

A concern with many previous QTL mapping studies is that they employ mapping populations initiated with just two strains, and use individuals subjected to very few rounds of meiotic recombination, limiting the scope of the genetic variation interrogated, and limiting the mapping resolution achievable (e.g., [26]). Here, we employ the DSPR (*Drosophila* Synthetic Population Resource [27, 28]) - a multiparental, advanced intercross panel of RILs (Recombinant Inbred Lines) - to dissect genetic variation in lifespan in mated female *Drosophila*, resolving five modest-effect QTL to relatively short genomic regions (0.1-1.2Mb). We also use RNAseq to identify genes showing differential expression between young and old animals in a subset of DSPR lines. The set of genes exhibiting age-related changes in gene expression in our study show significant overlap with previous such studies in flies, and implicate small numbers of highly plausible aging candidate genes within mapped QTL. Some of the loci we highlight were already considered candidates to contribute to aging based on studies of induced mutations, for instance *Relish*, a gene known to be involved in immune response.

MATERIALS AND METHODS

Mapping population

The DSPR is a large panel of RILs derived from a multi-parental, advanced generation intercross [28]. Each of the two populations - pA and pB - was initiated from a set of eight, highly-inbred founders, and was maintained as a pair of independent subpopulations - pA1, pA2, pB1, and pB2 - for 50 generations. Subsequently ~800 RILs per population were established via 25 generations of full sib mating, and genotyped via Restriction site Associated DNA sequencing (RADseq).

Since all founder lines were also sequenced to 50X coverage, we were able to use a hidden Markov Model (HMM) to elucidate the mosaic founder structure of each RIL. Full details of the construction of the DSPR are presented in King et al. [28].

Lifespan assay

Briefly, our assay was conducted as follows: Each RIL was copied from our stock collection in a single vial, and in the next generation expanded to two replicate experimental vials. In all cases adults were cleared as needed to maintain roughly equal egg density across vials. Nine days after egg laying any emerged adults were cleared from experimental vials. After 48 hours, 0-2 day old flies were transferred to fresh media, and held for 24 hours to ensure mating. Subsequently, 30 mated 1-3 day old female flies per RIL were collected under CO_2 anesthesia into a single assay vial. Flies were transferred to fresh media every two days for the first two weeks of life, and every three days thereafter, and flies were scored daily until half the females were dead. We tracked vials and genotypes using systems of anonymous barcodes, a barcode reader, and custom R code (r-project.org) designed to record the number of dead flies each day, trace all anonymous barcodes back to the original RIL genotype, and find the median lifespan for females from each RIL assayed.

We collected median lifespan data for mated females from 805 pB DSPR RILs, testing each RIL in one of four experimental blocks. All fly rearing and maintenance was carried out at 25ºC, 50% relative humidity, a 12 hour light/12 hour dark cycle, and used cornmeal-molasses-yeast media in standard, narrow *Drosophila* vials.

QTL mapping

The analytical framework used to identify QTL in the DSPR is described in detail in King et al. [28], and the power and properties of the mapping approach is presented in King et al. [27]. Briefly, the HMM assigns to each region in each RIL a probability the genotype is one of 36 possible homo- or heterozygous states. Since the vast majority of the positions in the RILs are homozygous, we generate eight additive homozygous probabilities per position, and regress RIL median lifespan on these probabilities. Since we see variation over experimental blocks (Figure 2.S1) we additionally include "block" as a covariate. We note that because lines from the pB1 and pB2 subpopulations were segregated into different blocks for the lifespan assay, some of the block-to-block variation is likely due to differences between subpopulations in addition to technical, experimental variation.

QTL were identified as peaks reaching a 5% genomewide, permutation-derived threshold [29], and we used 2-LOD support intervals to put confidence intervals on the true positions of QTL [27]. All mapping was carried out using the DSPRqtl R package (github.com/egking/DSPRqtl; FlyRILs.org).

RNAseq

In the course of assaying lifespan we collected samples of young (1-3 days old) and old (median lifespan for genotype) females from a large number of RILs. Each experimental sample consisted of a group of 10 females of the same genotype collected under CO_2 anesthesia and snap-frozen using liquid nitrogen. For each sample to be used for RNAseq we removed heads from bodies

(thorax + abdomen) by vortexing tubes containing frozen female flies, separating heads and bodies with a paintbrush over a dry ice-cooled aluminum block. RNA was isolated from each tissue sample using TRIzol reagent (15596-018, ThermoFisher Scientific) following the manufacturers protocol, except that for head samples we scaled down all volumes to 1/4 the recommended amounts.

To examine expression in bodies we selected 10 RILs with a relatively short lifespan, and 10 with longer lifespan (Supplementary Data 2.1). Equal amounts of total RNA from each of the appropriate 10 samples were combined to generate four pools; short-lived/young, short-lived/old, long-lived/young, and long-lived/old. Each pool was then cleaned through an RNeasy Mini column (74104, Qiagen), used to generate a standard TruSeq RNAseq library (version 2, Illumina), and sequenced on an Illumina HiSeq 2500 instrument (KU Genome Sequencing Core) to generate single-end 100bp reads (see SRA accession SRP072382). Quality trimming via sickle (version 1.200, github.com/najoshi/sickle) resulted in 34.2-39.5 million reads per sample. We used TopHat (version 2.0.12, tophat.cbcb.umd.edu; [30, 31]) to assemble reads to the *D. melanogaster* reference genome (NCBI build 5.3, tophat.cbcb.umd.edu/igenomes.shtml), resulting in 84.0-87.1% reads aligning, and Cuffdiff (version 2.1.1, cufflinks.cbcb.umd.edu; [32-34]) to identify differentially expressed genes in four pairwise contrasts (short-lived/young *versus* short-lived/old, long-lived/young *versus* long-lived/old, short-lived/young *versus* long-lived/young, and short-lived/old *versus* long-lived/old). We consider a gene to be differentially expressed if it survives a genomewide, per contrast Benjamini-Hochberg 5% False Discovery Rate (FDR) correction for multiple testing.

To investigate expression in heads we selected six genotypes (Additional file 3), made RNAseq libraries for the six pairs of young and old head samples, and sequenced to generate

paired-end 50bp reads (see SRA accession SRP072396). Following quality trimming we had 14.1-26.0 million read pairs per sample, and genome alignment resulted in 78.8-90.9% reads mapping. Statistical testing was carried out to find genes differentially expressed (FDR = 5%) between the heads of young and old flies, treating the separate RIL genotypes as replicates.

RESULTS

Variation in lifespan in the DSPR

We observed substantial lifespan variation among the 805 DSPR RILs tested (Figure 2.1), with median mated female lifespan averaging 55.0 days, ranging from 16.4-80.6 days across RILs. Since each RIL was assayed in just one block, some fraction of this variation is due to technical variation across blocks (Figure 2.S1). Nonetheless, the scale of lifespan variation we see is remarkably similar to that observed in a screen of virgin females from 197 *Drosophila* Genetic Reference Panel, DGRP lines (mean = 55.3 days, range = 22.1-80.3 days; [35]).

Given the number of RILs tested, to streamline phenotype data collection we elected to score RILs for median lifespan, allowing us to discard assay vials at that point, and avoid waiting for all flies in a vial to die. Although our data collection pipeline did not allow the calculation of mean lifespan for each RIL, results from the DGRP show that the correlation between mean and median lifespan for a set of inbred lines is very strong ($r = 0.97$, $p < 10^{-15}$; [35]). One caveat with our use of a phenotype based on the median lifespan from a single replicate vial per genotype is that we are unable to estimate heritability for lifespan in the DSPR.

QTL for variation in lifespan

We mapped five QTL for lifespan in the DSPR (Figure 2.2, Table 2.1, Additional file 4: Table S2). Each QTL explains a modest fraction of the among-line variation for lifespan (4.0-5.2%, Table 2.1), and assuming the QTL are independent and act additively, collectively explain 22.2% of the genetic variation for lifespan in the DSPR. With 800 RILs the power to identify common biallelic or multiallelic QTL contributing 5% to the total variation in the RIL panel is 80-90% [27]. This implies that any undetected genetic factors contributing to lifespan variation in the DSPR either have small effects on variation, or are rare in the panel.

A feature of multi-parental mapping panels such as the DSPR is that we can estimate the effects of each founder allele at mapped QTL, and can determine those founders that are likely to harbor alleles contributing to long lifespan. Figure 2.3 shows the founder allele effects for all five mapped QTL. It is not obvious from this plot that loci contributing to lifespan variation generally segregate for two alleles (e.g., a "high" and a "low" allele), and instead may segregate for multiple alleles, each with different effects on phenotype. Of course, since our QTL are mapped to intervals containing multiple genes (Table 2.1) we cannot discount the possibility that mapped QTL are due to the action of multiple genes. Regardless, it is possible to identify pairs of founders that appear to harbor alleles with contrasting effects on lifespan. For example, RILs carrying genetic material from founders B5 and B6 at Q2 have relatively low, and relatively high lifespan, respectively (Figure 2.3). Genetic differences between these founders in the Q2 interval are likely to be enriched for variants causally contributing to lifespan.

The five QTL are mapped to regions encompassing 660kb (Q1), 660kb (Q2), 510kb (Q3), 1.2Mb (Q4), and 80kb (Q5) of the *D. melanogaster* genome (Table 2.1). The Q4 interval is relatively large since this QTL resides near the chromosome 3 centromere where recombination is suppressed. Aside from Q4, QTL intervals include 11-93 protein-coding genes (Table 2.1). To determine whether any of the genes encompassed by mapped QTL have previously been implicated in aging and/or lifespan regulation, we searched FlyBase [36] to identify genes tagged with controlled vocabularies that included the words "aging", "lifespan", "lived", and "longevity" (Supplementary Data 2.2) We identified a total of 568 candidate genes, 14 of which reside within QTL intervals (Table 2. 2).

Comparison with previous mapping studies

Candidate aging genes extracted from FlyBase are often associated with longevity based on mutant phenotypes (e.g., *Cbs*, [37]), and may or may not harbor naturally-segregating variation affecting lifespan. Thus, we sought to compare our data to previous studies mapping lifespan loci among naturally-derived chromosomes. A number of previous studies have used various mapping designs to identify QTL contributing to variation in lifespan and aging in *D. melanogaster* [26, 38-49], and all five of the QTL we map in the DSPR overlap with at least one QTL mapped in a prior study (Figure 2.S2). While this observation gives some additional confidence in our phenotype and mapping, we note that the 13 studies we highlight mapped well over 100 QTL, and mapped intervals that collectively implicate 93.4% of the *D. melanogaster* genome (Figure 2.S2). This phenomenon of aging QTL implicating large fractions of the *Drosophila* genome has been noted previously [39]. Using a resampling procedure we tested how often five non-overlapping, randomly-positioned QTL of the same physical size as the set

mapped in this study overlapped previously identified QTL; Over 1000 runs, 85% of the time each of the five simulated QTL overlap at least one QTL mapped in a prior study, implying the overlap we see in our real data is expected. While the complexity of the genetic architecture of the phenotype may go some way to explaining the observation that aging QTL blanket the genome, lack of resolution in QTL mapping studies using animals that have passed through a small number of generations of meiotic recombination is likely also a major factor.

A more high-resolution study was conducted by Burke et al. [50]. Using animals derived the highly-recombinant "synthetic" 8-way populations from which the DSPR was derived, they compared allele frequencies in extremely old cohorts of females to those from randomly-selected, control females. Across all replicate populations eight regions survived a 5% false positive rate, and none of these overlap with the QTL we map here. Considering those additional eight regions from Burke et al. [50] surviving a more liberal 50% false positive rate threshold, one overlaps with our Q4 at the chromosome 3 centromere.

Ivanov et al. [35] recently used the DGRP to carry out a genomewide association study for lifespan using virgin females. Although no variant in the SNP-based GWAS, and no gene in the gene-based GWAS, survived a correction for multiple testing, likely due to the low power of the DGRP design [51], a number of variants and genes showed nominally-significant association tests. Such tests may be enriched for true causative variants/genes. Of the 50 SNP association tests with the lowest *P*-values, just one is within a region implicated by a QTL mapped in this study, a variant present within the *bves* gene [35] that is within our Q3. Although there is no specific information regarding the effect of *bves* on lifespan in FlyBase [36], an insertion mutation in the gene has been shown to increase the susceptibility of *Drosophila* to the fungal

pathogen, *Metarhizium anisopliae* [52]. None of the top-ranked gene-based DGRP burden-type tests carried out by Ivanov et al. [35] fall within our QTL intervals.

Regulatory candidate genes for lifespan

It is likely that some fraction of the sites that contribute to among-individual variation in a complex phenotype are regulatory in origin [53, 54]. Thus, we used RNAseq to identify genes differentially expressed between young and old female flies in heads and bodies, and between long-lived and short-lived genotypes for bodies only. Such candidate genes may plausibly harbor functional regulatory variants impacting lifespan.

For the body RNAseq we extracted RNA from samples of young and old flies from ten long-lived and ten short-lived RILs, mixed RNA to generate four pools each containing material from ten samples, generated and sequenced four libraries, and tested for differential gene expression in four pairwise contrasts: short-lived/young *versus* short-lived/old, long-lived/young *versus* long-lived/old, short-lived/young *versus* long-lived/young, and short-lived/old *versus* long-lived/old. After analysis we identified 155 genes differentially expressed between young and old flies in short-lived genotypes (22 down with age, 133 up with age), and 160 differentially expressed between young and old flies in long-lived genotypes (83 down with age, 77 up with age). Sixty-six genes overlap between these two sets, and all 66 show the same direction of age-related expression change in short- and long-lived animals, implying consistency in the pattern of age-related gene expression change across genotypes. We additionally identified 9 (16) genes showing significantly different expression in young (old) females when comparing short- and

long-lived genotypes. Overall, 252 genes survive a genomewide FDR threshold of 5% in at least one contrast.

For the head-specific RNAseq we extracted RNA from samples of young and old flies from six RILs, generated and sequenced separate libraries for each of the 12 samples, and identified 1,940 genes differentially expressed between young and old flies in heads (995 down with age, 945 up with age; Additional file 10: Table S6). Given that separate RILs were treated as replicates in the head RNAseq analysis, and assuming some consistency in the age-related patterns of expression across RILs, our power to detect small changes in expression in this head analysis is likely higher than for the body analysis that lacks replication at this level. Nonetheless, there was significant overlap - 130 genes - between the set of 249 genes showing differential expression between young and old flies in bodies, and the set of 1,940 showing expression differences between young and old flies in heads (Fisher's Exact Test, $p < 10^{-15}$, assuming 14,000 genes in the *D. melanogaster* genome). Nearly all - 127/130 - of the genes in this overlapping set show expression changes in the same direction in bodies and heads. Thus, despite experimental and analytical differences, we find similarity in the age-related patterns of expression across tissues.

Employing the Gene Ontology, GO (geneontology.org; [55, 56]) we classified genes showing differential expression by function and their involvement in particular biological processes (see Supplementary Data 2.3 for a summary). In both the head and body datasets considered separately we found a significant enrichment of genes involved in defense and response to bacteria, recapitulating previous results [57]. We additionally found an enrichment of genes involved in egg coat formation in the body data only, finding 5/14 such genes, all of which decrease in expression with age, clearly associated with reduced reproductive output in older

females (see also [57]). Finally, in bodies we saw an enrichment of myofibril assembly genes (10/40 genes found, all of which decrease in expression with age), and in heads an enrichment of genes involved in the electron transport chain (42/86 found, and 39/42 go down with age), both observations potentially reflecting a general loss of vigor with age. Studies in both mice and humans have also shown that many components of the electron transport chain show reduced expression with age [58].

Several other groups have previously used array-based expression profiling to identify genes that change with age in various *D. melanogaster* populations. We sought to compare the results of our study with this other work, and determine the extent of overlap in the genes identified among experiments. We extracted information on genes showing age-related changes in expression from Pletcher et al. [57], Landis et al. [59], Lai et al. [60], Zhan et al. [61], and Carlson et al. [62], converted all gene names to the most current FlyBase gene IDs (see Supplementary Data 2.4), and examined for the number of overlapping genes. Overall, 83% of the genes we identify as differentially expressed in bodies were identified in at least one other study, and 59% of the genes we identify in heads replicated (Figure 2.S3). We assessed the statistical significance of overlap in the sets of genes identified using the R software package *SuperExactTest* [63] that can calculate the probability of intersection among any number of gene sets. Considering our head (252 genes) and body (1,940 genes) datasets separately, and assuming 14,000 total genes in the *Drosophila* genome, the number of genes that intersect between our study and three or more other sets of age-related genes is highly significant (all p-values $< 3.7 \times 10^{-15}$). Thus, while there are an array of biological and technical differences among studies, a core set of genes appear to be consistently identified as showing age-related changes in gene expression.

Of interest is whether any of the genes we identify in our RNAseq screen are present within genomic intervals implicated by mapped QTL. A total of 55 RNAseq candidates are present within these intervals; Two were identified only in our body experiment, 52 only in our head experiment, and one was observed in both studies. In all cases these genes were identified as differentially expressed with age, and none were found to be differentially expressed between short- and long-lived genotypes. Thirty-one of the 55 genes have been shown to have age-related changes in expression in previous studies, and 4/55 represent aging candidate genes identified in FlyBase (Table 2. 2); *dome*, *Ubqn*, and *Zw* (all under Q2) and *Rel* (under Q4), all of which show increased expression in the heads of older females. These genes present excellent candidates to harbor regulatory variation affecting lifespan.

DISCUSSION

We carried out an unbiased screen to identify loci segregating for allelic variation influencing lifespan of mated female *D. melanogaster*. By virtue of employing a multiparental advanced intercross population we were able to map putative aging genes to relatively small regions of the genome averaging 640kb (Table 2.1), aiding future resolution of the actual causative loci. We uncovered three X-linked and two autosomal QTL that collectively explain 22.3% of the among-genotype variation in lifespan in the DSPR (Table 2.1). We were unable to estimate the heritability for lifespan directly in the DSPR, since our measure of lifespan for each of the 805 RILs assayed is derived from the median time of death of a single cohort of 30 flies. Nonetheless, previous estimates of broad-sense heritability of lifespan in *Drosophila* are 0.32-

0.41 (Khazaeli et al., 2005; Ivanov et al., 2015), suggesting that the QTL we identify likely explain very small fractions of the total phenotypic variation for lifespan.

Resolving candidates contributing to natural variation in aging

A benefit of mapping with high resolution in an advanced intercross population is that modest numbers of genes are implicated, allowing plausible candidates to be highlighted for future experimental tests. Below we summarize those plausibly functional loci residing within each of our mapped QTL.

Q1 (14A6-15A3) overlaps with lifespan QTL identified in studies by Reiwitch & Nuzhdin [46] and Defays et al. [39], and several of the 84 genes implicated by Q1 have been previously implicated in aging in flies (Table 2.2). A *caz* deletion mutation exhibited reduced longevity in comparison to wildtype [64], as did a *hang* P-element insertion mutation [65]. In addition, copy number at the *meiotic 41* gene has been shown to affect lifespan [66]. The gene *methuselah-like 1* (*mthl1*) is annotated in FlyBase as being involved in the determination of adult lifespan [36], although this appears to be entirely due to the sequence similarity of this gene to *methuselah*, a classic aging candidate gene [67]. We also identified 14 genes that change in expression between young and old flies in the head. Notably *UQCR-14*, which appears to be involved in mitochondrial electron transport [68], shows decreased expression with age in our study, reduced expression with age in whole females in both regular food and caloric restriction conditions in Pletcher et al. [57], lower expression with age in whole males [59], and changes expression with age in brain-tissue derived from males [61].

Q2 (18C8-19C1) was found in the same position as QTL mapped in three previous studies [39, 46, 47], although the QTL we map is considerably smaller in size, implicating 93 protein-coding genes. Several strong aging candidate genes are present in this interval (Table 2.2). A point mutation in *car* shows significantly reduced lifespan in males [69], RNAi knockdown of the mitochondrial electron transport chain complex IV component gene *CG18809* leads to a 16-19% increase in lifespan in female flies [70], a dominant negative version of *dome* increases mortality in a *G9a* mutant background [71], silencing *Ubqn* in the nervous system shortens lifespan in males and leads to neurodegeneration [72], and overexpression of *Zw* (glucose-6-phosphate dehydrogenase) increases lifespan [73]. *dome*, *Ubqn*, and *Zw* are also among the genes we identified as differentially expressed in heads between young and old animals, and these three genes all show enhanced expression with age.

Q3 (19E4-20A1) resides close to Q2 (Figure 2.2), however the 2-LOD drop confidence intervals of the peaks do not overlap (Table 2.1), and the founder allele effect plots show different patterns (Figure 2.3), so we can be reasonably confident the QTL represent separate loci. The positions of our Q1, Q2, and Q3 all overlap one of the broad QTL mapped by Defays et al. [39], highlighting the resolution of our study. Two *a priori* aging candidate genes are present within the Q3 interval (Table 2.2); *Cbs* overexpression leads to increased lifespan [37], and *Mgstl* null mutants exhibit reduced lifespan compared to wildtype controls [74].

Q4 (84F1-85D11) is the broadest peak we map, implicating 155 genes, likely because the QTL resides close to the chromosome 3 centromere, a site of reduced crossover rate. Our QTL overlaps loci previously mapped in five studies [39, 40, 43, 47, 49], although the region we implicate is substantially smaller than in most of these studies. Several genes in the Q4 interval have been previously implicated in *Drosophila* longevity (Table 2.2). *Coq2* is involved in the

synthesis of Coenzyme Q (ubiquinone; [75]), an essential electron carrier in the mitochondrial electron transport chain. Heterozygous genotypes with just one functional copy of *Coq2* show lifespan extension in both males and females [75]. Genotypes with nonfunctional *Ibf2* are short-lived [76], there is some evidence for a slight reduction in lifespan in genotypes carrying a mutant for *pum* [77], and loss of function mutations in *Rel* - a gene critical in the induction of the immune response in flies - dramatically reduce survival time compared to controls [78]. *Rel* is also an excellent expression candidate for a role in lifespan regulation, since we found it to be increased in expression with age in heads, and three previous studies showed also showed increased *Rel* expression in older flies [57, 59, 60]. Q4 also harbors *polychaetoid*, the only gene identified in a P-element screen for lifespan extension mutations that overlapped our five QTL intervals [21]. A number of genes within Q4 show expression variation between young and old animals in our study. This set includes *CG8032*, which is also the only member of a set of 39 lifespan-reducing loci identified in a gain-of-function screen that is implicated by QTL mapped in the present study [22], and *Nmdmc*, overexpression of which has been shown to extend lifespan in flies [79]. Given the number of genes within Q4, and the ample evidence of multiple candidates present in the region, it is not unlikely that more than one gene in the region is responsible for the QTL we map.

Finally, we mapped Q5 (98E2-98E5) to a small interval on chromosome 3R containing just 11 genes (Table 2.1, Figure 2.2). This region has previously been implicated in the control of lifespan [44-46], although no strong *a priori* candidates are present. One of the loci within the Q5 interval, *wdn*, shows an age-related increase in expression in heads in our study, although this result was not recapitulated in any of the five other expression datasets we examined.

Replication among studies mapping naturally-segregating aging variants

Each of the five QTL we isolated in the DSPR co-localizes with the positions of QTL mapped for lifespan in at least one of the 13 other studies we examined (Figure 2.S2). It is clear from examining overlap among all studies that there is some commonality in the genomic regions implicated in the control of natural variation in aging. However, it is equally clear based on the lack of the overlap among studies with the highest level of resolution (this study along with [42, 45, 49]) that there are significant differences in the sets of loci implicated in different works (Figure 2.S2). Studies routinely employ different starting sets of genotypes, so at least some of the differences observed must be due to different mapping panels segregating for different subsets of functional allelic variation. However, differences in power among studies are also likely to play an important role in the differing results. It is most likely that aging is a highly polygenic trait, and that individual variants each underlie only a tiny fraction of lifespan variation, as evidenced by the small effects of the two genes replicated in multiple human GWAS for aging, *APOE* and *FOXO3A* [80]. If variant effects are routinely this low, even studies with reasonable sample sizes are likely to be underpowered; For instance, this study used 805 RILs, and has ~30% power to identify QTL contributing 2.5% to among-line variation in phenotype [27]. Thus, if the genetic architecture of lifespan is constructed from the effects of many, very small-effect variants, any given genomewide study may only find a small subset of the loci segregating for age-related variation.

A further important difference among studies in *Drosophila* is that the assays used to measure lifespan are frequently different, both at gross levels (e.g., studies may focus on different sexes), and at more subtle levels, such as any number of technical differences in the laboratory environments used to rear flies and maintain aging populations (e.g., temperature,

media composition, larval density). Loci contributing to aging have been shown to be sex-specific in many cases [26, 44], whether the flies are mated or virgin can alter the QTL identified [46], and many QTL have been shown to be highly environment-dependent [43, 47]. If the effects of functional alleles at aging genes are typically, or even often sensitive to the environment in this way, i.e., exhibit genotype by environment interaction [81-83], we should expect to routinely identify different sets of variants in different studies, with variants only being identified under those conditions under which they have detectable effects on phenotype. A key benefit of a consistent, chemically-defined diet for flies [84] would be to help minimize lab-to-lab variation in studies of life history traits, help enhance replicability of genetic effects across studies, and promote understanding of the mechanisms by which allelic variation leads to variation in aging under a single set of conditions.

Prospects

Regardless of the precise set of aging loci identified in mapping populations of *Drosophila*, there is clearly some consistency across studies in the pathways implicated in the aging process. This is most easily seen in the various expression profiling experiments that have been carried out, where core groups of genes robustly and consistently show age-related changes in expression, notably antimicrobial defense response genes that are routinely upregulated during aging, and genes involved in the electron transport chain that are routinely downregulated with age. Thus, there is hope that the genes implicated by QTL mapping studies, regardless of their differences across studies, could provide valuable inroads into a mechanistic understanding of the pathways involved in aging. In this regard, our identification of *UQCR-14*, a gene within Q1 that is involved in electron transport and shows a decrease in expression with age, *CG18809*, a gene

within Q2 that encodes a component of the electron transport chain, and *Relish*, a gene under Q4 that is involved in mobilizing the antimicrobial response, and shows increased expression in aged animals, represent excellent candidates for future functional analysis, and to identify causative sequence-level variation underlying aging. The prospects for direct functional validation of age-related variation in model systems via allele swapping - moving "high" alleles into "low" backgrounds and *vice versa* - using CRISPR-Cas9 editing are strong, and will obviate the need to "validate" natural allelic effects with synthetic constructs (e.g., RNAi). The ability to examine whole organism phenotypes, in addition to cellular and physiological phenomena, in specifically edited animal models is a considerable strength of model organisms that will allow the exploration of aging pathways that may also be implicated in humans.

Authors' contributions

Conceived and designed the experiment: CAH, SJM. Collected data: CAH, GAR. Analyzed data: CAH, GAR, SJM. Wrote the paper: CAH, SJM. All authors edited, read, and approved the final manuscript.

Acknowledgements

We thank Xiaofei Wang, K-INBRE Bioinformatics Core, for help with RNAseq analysis, and Rob Unckless for comments on an earlier version of the manuscript. Conceived and designed the experiment: CAH, SJM. Collected data: CAH, GAR. Analyzed data: CAH, GAR, SJM. Wrote the paper: CAH, SJM. All authors edited, read, and approved the final manuscript.

REFERENCES

1. Arias E: **United States life tables, 2011**. In: *National vital statistics reports.* vol. 64. Hyattsville, MD: National Center for Health Statistics; 2015.

2. Niccoli T, Partridge L: **Ageing as a risk factor for disease**. *Curr Biol* 2012, **22**(17):R741-752.

3. Hjelmborg JB, Iachine I, Skytthe A, Vaupel JW, McGue M, Koskenvuo M, Kaprio J, Pedersen NL, Christensen K: **Genetic influence on human lifespan and longevity**. *Hum Genet* 2006, **119**(3):312-321.

4. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW: **The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900**. *Hum Genet* 1996, **97**(3):319-323.

5. McGue M, Vaupel JW, Holm N, Harvald B: **Longevity is moderately heritable in a sample of Danish twins born 1870-1880**. *J Gerontol* 1993, **48**(6):B237-244.

6. Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, Kamatani Y, Bennet AM, Tamm R, Trompet S *et al*: **Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age**. *Hum Mol Genet* 2014, **23**(16):4420-4432.

7. Sebastiani P, Solovieff N, Dewan AT, Walsh KM, Puca A, Hartley SW, Melista E, Andersen S, Dworkis DA, Wilk JB *et al*: **Genetic signatures of exceptional longevity in humans**. *PLoS One* 2012, **7**(1):e29848.

8. Zeng Y, Nie C, Min J, Liu X, Li M, Chen H, Xu H, Wang M, Ni T, Li Y *et al*: **Novel loci and pathways significantly associated with longevity**. *Sci Rep* 2016, **6**:21243.

9. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, DeStafano AL, Bis JC, Beecham GW, Grenier-Boley B *et al*: **Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease**. *Nat Genet* 2013, **45**(12):1452-1458.

10. Tan Q, Zhao JH, Zhang D, Kruse TA, Christensen K: **Power for genetic association study of human longevity using the case-control design**. *Am J Epidemiol* 2008, **168**(8):890-896.

11. Pritchard JK: **Are rare variants responsible for susceptibility to complex diseases?** *Am J Hum Genet* 2001, **69**(1):124-137.

12. Spencer CC, Su Z, Donnelly P, Marchini J: **Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip**. *PLoS Genet* 2009, **5**(5):e1000477.

13. Thornton KR, Foran AJ, Long AD: **Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect**. *PLoS Genet* 2013, **9**(2):e1003258.

14. Fontana L, Partridge L: **Promoting health and longevity through diet: from model organisms to humans**. *Cell* 2015, **161**(1):106-118.

15. Fontana L, Villareal DT, Weiss EP, Racette SB, Steger-May K, Klein S, Holloszy JO, Washington University School of Medicine CG: **Calorie restriction or exercise: effects on coronary heart disease risk factors. A randomized, controlled trial**. *Am J Physiol Endocrinol Metab* 2007, **293**(1):E197-202.

16. Yang L, Licastro D, Cava E, Veronese N, Spelta F, Rizza W, Bertozzi B, Villareal DT, Hotamisligil GS, Holloszy JO *et al*: **Long-Term Calorie Restriction Enhances Cellular Quality-Control Processes in Human Skeletal Muscle**. *Cell Rep* 2016, **14**(3):422-428.

17. Kenyon C, Chang J, Gensch E, Rudner A, Tabtiang R: **A C. elegans mutant that lives twice as long as wild type**. *Nature* 1993, **366**(6454):461-464.

18. Kimura KD, Tissenbaum HA, Liu Y, Ruvkun G: **daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans**. *Science* 1997, **277**(5328):942-946.

19. Clancy DJ, Gems D, Harshman LG, Oldham S, Stocker H, Hafen E, Leevers SJ, Partridge L: **Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein**. *Science* 2001, **292**(5514):104-106.

20. Selman C, Lingard S, Choudhury AI, Batterham RL, Claret M, Clements M, Ramadani F, Okkenhaug K, Schuster E, Blanc E *et al*: **Evidence for lifespan extension and delayed age-related biomarkers in insulin receptor substrate 1 null mice**. *FASEB J* 2008, **22**(3):807-818.

21. Magwire MM, Yamamoto A, Carbone MA, Roshina NV, Symonenko AV, Pasyukova EG, Morozova TV, Mackay TF: **Quantitative and molecular genetic analyses of mutations increasing Drosophila life span**. *PLoS Genet* 2010, **6**(7):e1001037.

22. Nakayama M, Ishibashi T, Ishikawa HO, Sato H, Usui T, Okuda T, Yashiro H, Ishikawa H, Taikou Y, Minami A *et al*: **A gain-of-function screen to identify genes that reduce lifespan in the adult of Drosophila melanogaster**. *BMC Genet* 2014, **15**:46.

23. Paaby AB, Schmidt PS: **Dissecting the genetics of longevity in Drosophila melanogaster**. *Fly (Austin)* 2009, **3**(1):29-38.

24. Mackay TF: **The nature of quantitative genetic variation for Drosophila longevity**. *Mech Ageing Dev* 2002, **123**(2-3):95-104.

25. De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG, Mackay TF: **Dopa decarboxylase (Ddc) affects variation in Drosophila longevity**. *Nat Genet* 2003, **34**(4):429-433.

26. Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF: **Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster**. *Proc Natl Acad Sci U S A* 1997, **94**(18):9734-9739.

27. King EG, Macdonald SJ, Long AD: **Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits**. *Genetics* 2012, **191**(3):935-949.

28. King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, Long AD, Macdonald SJ: **Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource**. *Genome Res* 2012, **22**(8):1558-1566.

29. Churchill GA, Doerge RW: **Empirical threshold values for quantitative trait mapping**. *Genetics* 1994, **138**(3):963-971.

30. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions**. *Genome Biol* 2013, **14**(4):R36.

31. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq**. *Bioinformatics* 2009, **25**(9):1105-1111.

32. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: **Improving RNA-Seq expression estimates by correcting for fragment bias**. *Genome Biol* 2011, **12**(3):R22.

33. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L: **Differential analysis of gene regulation at transcript resolution with RNA-seq**. *Nat Biotechnol* 2013, **31**(1):46-53.

34. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L: **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation**. *Nat Biotechnol* 2010, **28**(5):511-515.

35. Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TF, Partridge L, Thornton JM: **Longevity GWAS Using the Drosophila Genetic Reference Panel**. *J Gerontol A Biol Sci Med Sci* 2015, **70**(12):1470-1478.

36. Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ, FlyBase c: **FlyBase: establishing a Gene Group resource for Drosophila melanogaster**. *Nucleic Acids Res* 2016, **44**(D1):D786-792.

37. Kabil H, Kabil O, Banerjee R, Harshman LG, Pletcher SD: **Increased transsulfuration mediates longevity and dietary restriction in Drosophila**. *Proc Natl Acad Sci U S A* 2011, **108**(40):16831-16836.

38. Curtsinger JW, Khazaeli AA: **Lifespan, QTLs, age-specificity, and pleiotropy in Drosophila**. *Mech Ageing Dev* 2002, **123**(2-3):81-93.

39. Defays R, Gomez FH, Sambucetti P, Scannapieco AC, Loeschcke V, Norry FM: **Quantitative trait loci for longevity in heat-stressed Drosophila melanogaster**. *Exp Gerontol* 2011, **46**(10):819-826.

40. Forbes SN, Valenzuela RK, Keim P, Service PM: **Quantitative trait loci affecting life span in replicated populations of Drosophila melanogaster. I. Composite interval mapping**. *Genetics* 2004, **168**(1):301-311.

41. Geiger-Thornsberry GL, Mackay TF: **Quantitative trait loci affecting natural variation in Drosophila longevity**. *Mech Ageing Dev* 2004, **125**(3):179-189.

42. Lai CQ, Leips J, Zou W, Roberts JF, Wollenberg KR, Parnell LD, Zeng ZB, Ordovas JM, Mackay TF: **Speed-mapping quantitative trait loci using microarrays**. *Nat Methods* 2007, **4**(10):839-841.

43. Leips J, Mackay TF: **Quantitative trait loci for life span in Drosophila melanogaster: interactions with genetic background and larval density**. *Genetics* 2000, **155**(4):1773-1788.

44. Leips J, Mackay TF: **The complex genetic architecture of Drosophila life span**. *Exp Aging Res* 2002, **28**(4):361-390.

45. Nuzhdin SV, Khazaeli AA, Curtsinger JW: **Survival analysis of life span quantitative trait loci in Drosophila melanogaster**. *Genetics* 2005, **170**(2):719-731.

46. Reiwitch SG, Nuzhdin SV: **Quantitative trait loci for lifespan of mated Drosophila melanogaster affect both sexes**. *Genet Res* 2002, **80**(3):225-230.

47. Vieira C, Pasyukova EG, Zeng ZB, Hackett JB, Lyman RF, Mackay TF: **Genotype-environment interaction for quantitative trait loci affecting life span in Drosophila melanogaster**. *Genetics* 2000, **154**(1):213-227.

48. Wang MH, Lazebny O, Harshman LG, Nuzhdin SV: **Environment-dependent survival of Drosophila melanogaster: a quantitative genetic analysis**. *Aging Cell* 2004, **3**(3):133-140.

49. Wilson RH, Morgan TJ, Mackay TF: **High-resolution mapping of quantitative trait loci affecting increased life span in Drosophila melanogaster**. *Genetics* 2006, **173**(3):1455-1463.

50. Burke MK, King EG, Shahrestani P, Rose MR, Long AD: **Genome-wide association study of extreme longevity in Drosophila melanogaster**. *Genome Biol Evol* 2014, **6**(1):1-11.

51. Long AD, Macdonald SJ, King EG: **Dissecting complex traits using the Drosophila Synthetic Population Resource**. *Trends Genet* 2014, **30**(11):488-495.

52. Lu HL, Wang JB, Brown MA, Euerle C, St Leger RJ: **Identification of Drosophila Mutants Affecting Defense to an Entomopathogenic Fungus**. *Sci Rep* 2015, **5**:12350.

53. Gilad Y, Rifkin SA, Pritchard JK: **Revealing the architecture of gene regulation: the promise of eQTL studies**. *Trends Genet* 2008, **24**(8):408-415.

54. Torres JM, Gamazon ER, Parra EJ, Below JE, Valladares-Salgado A, Wacher N, Cruz M, Hanis CL, Cox NJ: **Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait**. *Am J Hum Genet* 2014, **95**(5):521-534.

55. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**(1):25-29.

56. Consortium GO: **Gene Ontology Consortium: going forward**. *Nucleic Acids Res* 2015, **43**(Database issue):D1049-1056.

57. Pletcher SD, Macdonald SJ, Marguerie R, Certa U, Stearns SC, Goldstein DB, Partridge L: **Genome-wide transcript profiles in aging and calorically restricted Drosophila melanogaster**. *Curr Biol* 2002, **12**(9):712-723.

58. Zahn JM, Sonu R, Vogel H, Crane E, Mazan-Mamczarz K, Rabkin R, Davis RW, Becker KG, Owen AB, Kim SK: **Transcriptional profiling of aging in human muscle reveals a common aging signature**. *PLoS Genet* 2006, **2**(7):e115.

59. Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, Tavare S, Tower J: **Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster**. *Proc Natl Acad Sci U S A* 2004, **101**(20):7663-7668.

60. Lai CQ, Parnell LD, Lyman RF, Ordovas JM, Mackay TF: **Candidate genes affecting Drosophila life span identified by integrating microarray gene expression analysis and QTL mapping**. *Mech Ageing Dev* 2007, **128**(3):237-249.

61. Zhan M, Yamaza H, Sun Y, Sinclair J, Li H, Zou S: **Temporal and spatial transcriptional profiles of aging in Drosophila melanogaster**. *Genome Res* 2007, **17**(8):1236-1243.

62. Carlson KA, Gardner K, Pashaj A, Carlson DJ, Yu F, Eudy JD, Zhang C, Harshman LG: **Genome-Wide Gene Expression in relation to Age in Large Laboratory Cohorts of Drosophila melanogaster**. *Genet Res Int* 2015, **2015**:835624.

63. Wang M, Zhao Y, Zhang B: **Efficient Test and Visualization of Multi-Set Intersections**. *Sci Rep* 2015, **5**:16923.

64. Wang JW, Brent JR, Tomlinson A, Shneider NA, McCabe BD: **The ALS-associated proteins FUS and TDP-43 function together to affect Drosophila locomotion and life span**. *J Clin Invest* 2011, **121**(10):4118-4126.

65. Scholz H, Franz M, Heberlein U: **The hangover gene defines a stress pathway required for ethanol tolerance development**. *Nature* 2005, **436**(7052):845-847.

66. Symphorien S, Woodruff RC: **Effect of DNA repair on aging of transgenic Drosophila melanogaster: I. mei-41 locus**. *J Gerontol A Biol Sci Med Sci* 2003, **58**(9):B782-787.

67. Lin YJ, Seroude L, Benzer S: **Extended life-span and stress resistance in the Drosophila mutant methuselah**. *Science* 1998, **282**(5390):943-946.

68. Homem CC, Steinmann V, Burkard TR, Jais A, Esterbauer H, Knoblich JA: **Ecdysone and mediator change energy metabolism to terminate proliferation in Drosophila neural stem cells**. *Cell* 2014, **158**(4):874-888.

69. Simonsen A, Cumming RC, Lindmo K, Galaviz V, Cheng S, Rusten TE, Finley KD: **Genetic modifiers of the Drosophila blue cheese gene link defects in lysosomal transport with decreased life span and altered ubiquitinated-protein profiles**. *Genetics* 2007, **176**(2):1283-1297.

70. Copeland JM, Cho J, Lo T, Jr., Hur JH, Bahadorani S, Arabyan T, Rabie J, Soh J, Walker DW: **Extension of Drosophila life span by RNAi of the mitochondrial respiratory chain**. *Curr Biol* 2009, **19**(19):1591-1598.

71. Merkling SH, Bronkhorst AW, Kramer JM, Overheul GJ, Schenck A, Van Rij RP: **The epigenetic regulator G9a mediates tolerance to RNA virus infection in Drosophila**. *PLoS Pathog* 2015, **11**(4):e1004692.

72. Li A, Xie Z, Dong Y, McKay KM, McKee ML, Tanzi RE: **Isolation and characterization of the Drosophila ubiquilin ortholog dUbqln: in vivo interaction with early-onset Alzheimer disease genes**. *Hum Mol Genet* 2007, **16**(21):2626-2639.

73. Legan SK, Rebrin I, Mockett RJ, Radyuk SN, Klichko VI, Sohal RS, Orr WC: **Overexpression of glucose-6-phosphate dehydrogenase extends the life span of Drosophila melanogaster**. *J Biol Chem* 2008, **283**(47):32492-32499.

74. Toba G, Aigaki T: **Disruption of the microsomal glutathione S-transferase-like gene reduces life span of Drosophila melanogaster**. *Gene* 2000, **253**(2):179-187.

75. Liu J, Wu Q, He D, Ma T, Du L, Dui W, Guo X, Jiao R: **Drosophila sbo regulates lifespan through its function in the synthesis of coenzyme Q in vivo**. *J Genet Genomics* 2011, **38**(6):225-234.

76. Cuartero S, Fresan U, Reina O, Planet E, Espinas ML: **Ibf1 and Ibf2 are novel CP190-interacting proteins required for insulator function**. *EMBO J* 2014, **33**(6):637-647.

77. Landis G, Bhole D, Lu L, Tower J: **High-frequency generation of conditional mutations affecting Drosophila melanogaster development and life span**. *Genetics* 2001, **158**(3):1167-1176.

78. Petersen AJ, Katzenberger RJ, Wassarman DA: **The innate immune response transcription factor relish is necessary for neurodegeneration in a Drosophila model of ataxia-telangiectasia**. *Genetics* 2013, **194**(1):133-142.

79. Yu S, Jang Y, Paik D, Lee E, Park JJ: **Nmdmc overexpression extends Drosophila lifespan and reduces levels of mitochondrial reactive oxygen species**. *Biochem Biophys Res Commun* 2015, **465**(4):845-850.

80. Wheeler HE, Kim SK: **Genetics and genomics of human ageing**. *Philos T R Soc B* 2011, **366**(1561):43-50.

81. Piper MD, Bartke A: **Diet and aging**. *Cell Metab* 2008, **8**(2):99-104.

82. Spencer CC, Howell CE, Wright AR, Promislow DE: **Testing an 'aging gene' in long-lived drosophila strains: increased longevity depends on sex and genetic background**. *Aging Cell* 2003, **2**(2):123-130.

83. Zhu CT, Ingelmo P, Rand DM: **GxGxE for lifespan in Drosophila: mitochondrial, nuclear, and dietary interactions that modify longevity**. *PLoS Genet* 2014, **10**(5):e1004354.

84. Piper MD, Blanc E, Leitao-Goncalves R, Yang M, He X, Linford NJ, Hoddinott MP, Hopfen C, Soultoukis GA, Niemeyer C *et al*: **A holidic medium for Drosophila melanogaster**. *Nat Methods* 2014, **11**(1):100-105.

Figure 2.1

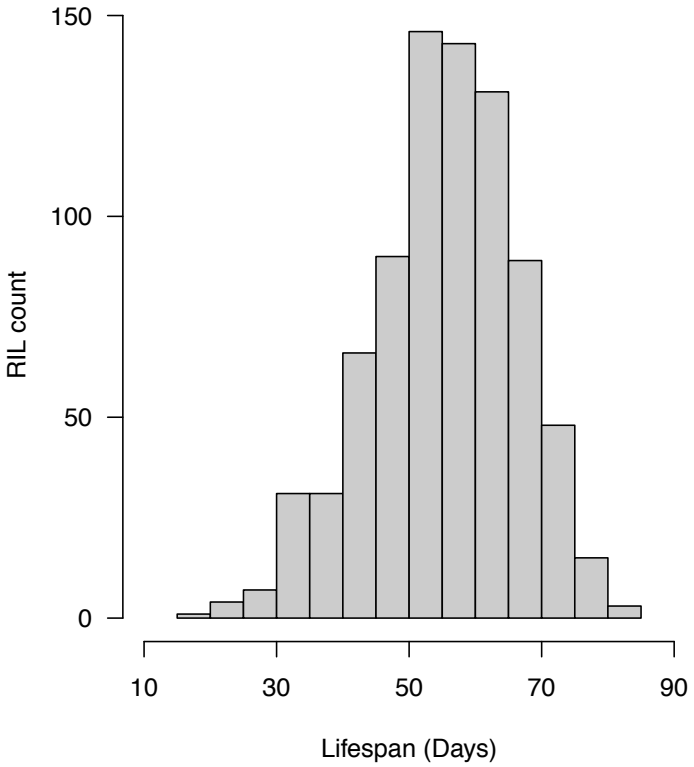


Figure 2.1 Distribution of female lifespan among DSPR RILs. We assayed lifespan for 805 RILs from the DSPR, measuring the phenotype as the time required for half the flies to die.

Figure 2.2

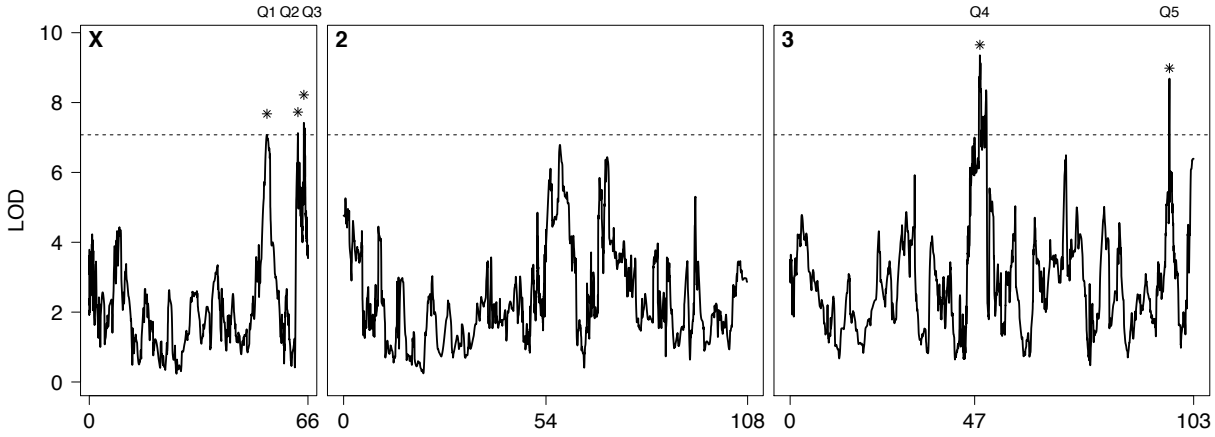


Figure 2.2 Genome scan for lifespan QTL. The black solid line indicates the LOD score following a scan for QTL contributing to variation in lifespan in the DSPR. The *x*-axis indicates genetic distance, and genetic positions 54 and 47 are the sites of the centromeres on chromosomes 2 and 3, respectively. The dashed line is a permutation-based genomewide 5% threshold (LOD = 7.08), and five QTL show peaks with LOD scores higher than this threshold. The positions of these QTL are indicated with asterisks, and their codes (Q1-Q5) are provided above the plot.

Figure 2.3

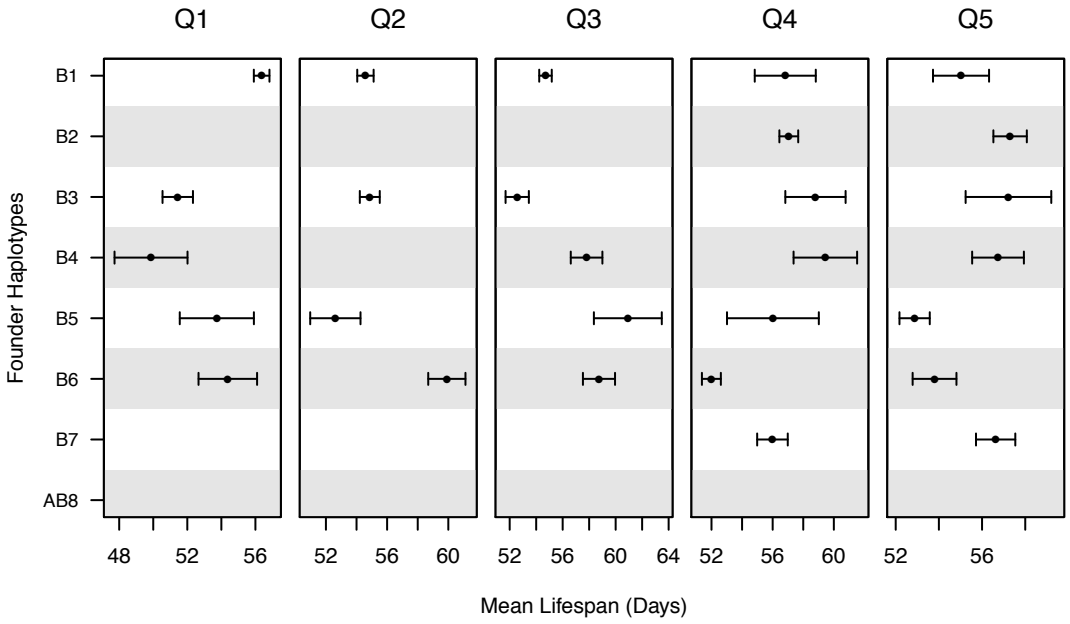

Q1
Q2
Q3
Q4
Q5
B1
B2
B3
B4
B5
B6
B7
AB8
Founder Haplotypes
48
52
56
52
56
60
52
56
60
64
52
56
60
52
56
Mean Lifespan (Days)

Figure 2.3 Founder allele strain effects at mapped lifespan QTL. Phenotype means (± 1 standard error) are presented for each founder at each QTL peak. Data is presented only for those founders present in at least 10 RILs at a probability > 0.95.

Table 2.1 Lifespan QTL mapped in the pB DSPR panel.

QTL	Peak LOD [a]	Chr. [b]	Physical Interval (Mb) [b]	Cytological Interval [b]	Number of Genes [c]	Variation explained [d]
Q1	7.1	X	16.0-16.7	14A6-15A3	84	4.0
Q2	7.1	X	19.5-20.2	18C8-19C1	93	4.0
Q3	7.4	X	20.9-21.4	19E4-20A1	51	4.2
Q4	9.4	3R	8.1-9.3	84F1-85D11	155	5.2
Q5	8.7	3R	28.7-28.8	98E2-98E5	11	4.9

[a] LOD score at the QTL peak.

[b] The chromosome arm on which the QTL resides, the physical position of the QTL interval (defined as a 2-LOD drop from the peak) in the *D. melanogaster* reference genome release 6, and the equivalent cytological interval.

[c] Number of protein-coding genes present within the QTL interval.

[d] The fraction of the among-line variation explained by the QTL.

Table 2.2 FlyBase aging candidate genes within mapped QTL.

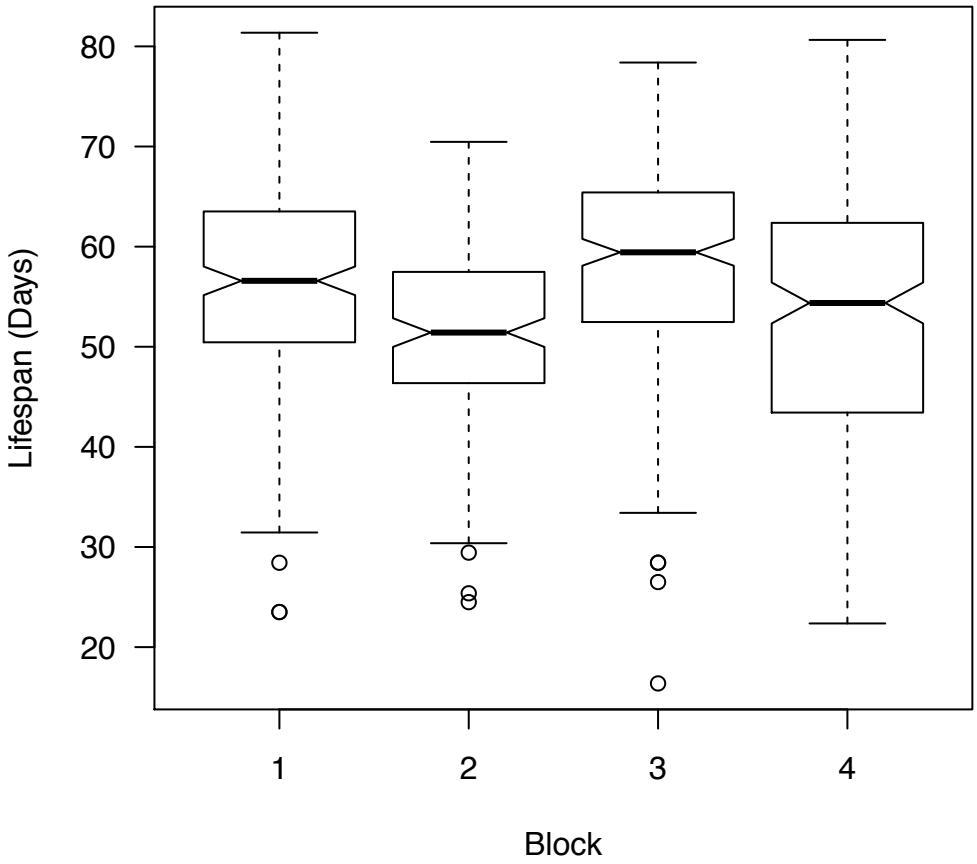
QTL [a]	Gene Name	Symbol
Q1 [b]	*cabeza*	*caz*
	hangover	*hang*
	methuselah-like 1	*mthl1*
Q2	*carnation*	*car*
	CG18809	–
	domeless	*dome* [c]
	Ubiquilin	*Ubqn* [c]
	Zwischenferment	*Zw* [c]
Q3	*Cystathionine beta-synthase*	*Cbs*
	Microsomal glutathione S-transferase-like	*Mgstl*
Q4	*Coenzyme Q biosynthesis protein 2*	*Coq2*
	Insulator binding factor 2	*Ibf2*
	pumilio	*pum*
	Relish	*Rel* [c]

[a] No genes from our FlyBase controlled vocabulary searches were present within the Q5 interval.

[b] The gene *CG32576*, which resides within Q1, was also tagged in our FlyBase search as "short lived" but this appears to be an annotation error [64].

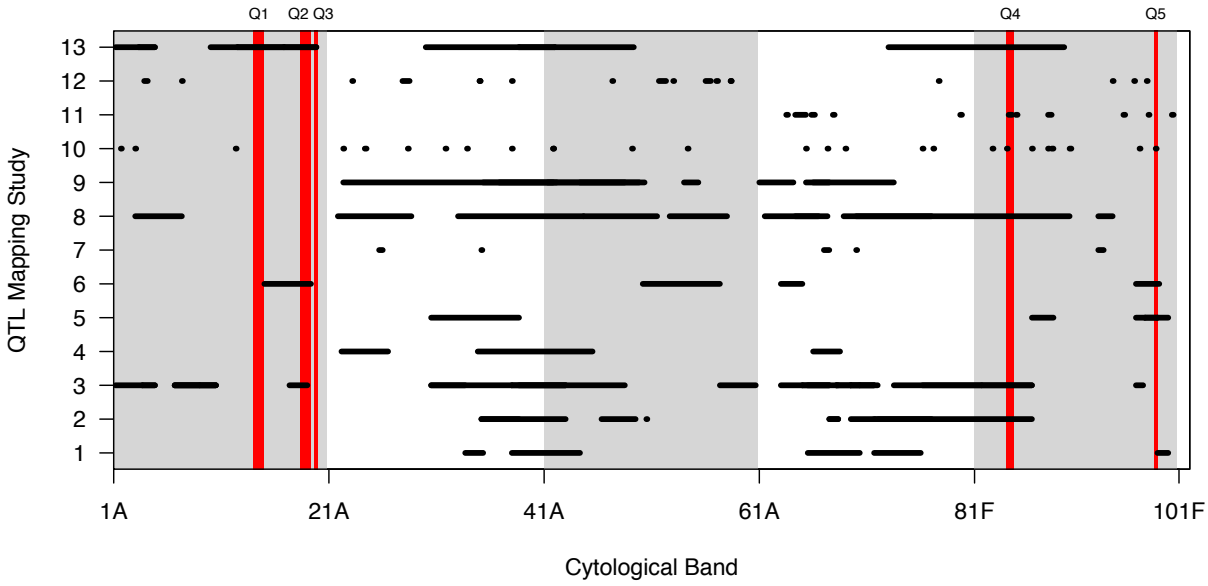
[c] These genes were also shown to increase in expression with age in female heads in our RNAseq study.

Figure 2.S1



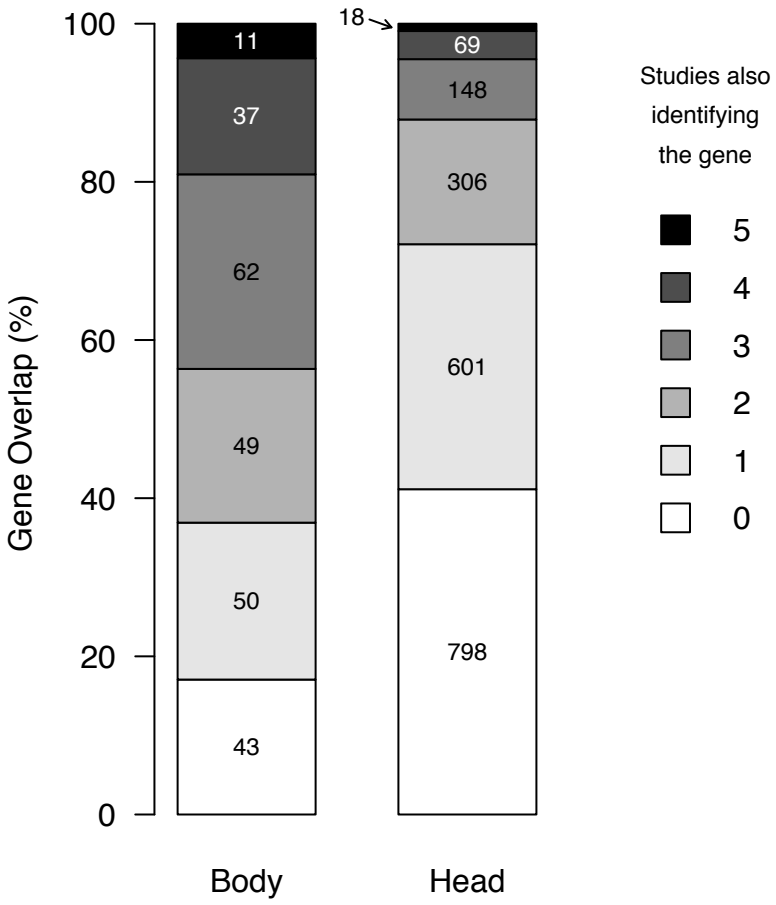
Supplementary Figure 2.1 Block-to-block variation in lifespan. We assayed 805 DSPR RILs for median female lifespan across four experimental blocks. There is significant variation across blocks in the average lifespan of the set of RILs assayed. The mean (± 1-SD) lifespan for each is 56.5 ± 9.93 (Block 1, 208 RILs), 50.5 ± 9.35 (Block 2, 150 RILs), 58.5 ± 10.42 (Block 3, 233 RILs), and 53.7 ± 11.98 (Block 4, 214 RILs). The figure was generated via the R "boxplot" command with standard settings.

Figure 2.S2


Q1
Q2 Q3
Q4
Q5
QTL Mapping Study
13
12
11
10
9
8
7
6
5
4
3
2
1
1A
21A
41A
61A
81F
101F
Cytological Band

Supplementary Figure 2.2 Lifespan QTL mapped in previous studies. We extracted the positions of QTL mapped for various lifespan/aging traits from 13 publications. All studies reported the positions of QTL intervals as cytological band locations, and the figure shows a horizontal black bar for each region implicated in each study. Alternating gray and white panels refer to the *D. melanogaster* chromosome arms (X, 2L, 2R, 3L, 3R, and 4). The positions of the five QTL we map in the DSPR are shown as vertical red bars. Studies listed along the y-axis are: (1) Nuzhdin et al., 1997 (PMID: 9275193), (2) Leips and Mackay, 2000 (PMID: 10924473), (3) Vieira et al., 2000 (PMID: 10628982), (4) Curtsinger and Khazaeli, 2002 (PMID: 11718803), (5) Leips and Mackay, 2002 (PMID: 12227919), (6) Reiwitch and Nuzhdin, 2002 (PMID: 12688661), (7) Geiger-Thornsberry and Mackay, 2004 (PMID: 15013662), (8) Forbes et al., 2004 (PMID: 15454544), (9) Wang et al., 2004 (PMID: 15153181), (10) Nuzhdin et al., 2005 (PMID: 15834144), (11) Wilson et al., 2006 (PMID: 16702433), (12) Lai et al., 2007 (PMID: 17873888), and (13) Defays et al., 2011 (PMID: 21798333). The positions of the QTL we map overlap with those mapped in studies 6 and 13 (Q1), 3, 6, and 13 (Q2), 13 (Q3), 2, 3, 8, 11, and 13 (Q4), and 5, 6, and 10 (Q5).

Figure 2.S3


100
80
60
40
20
0
Gene Overlap (%)
11
37
62
49
50
43
18
69
148
306
601
798
Body
Head
Studies also identifying the gene
5
4
3
2
1
0

Supplementary Figure 2.3 Overlap among expression candidates. We identified 252 genes showing differential expression in body tissue between young and old animals and/or between short- and long-lived genotypes. We additionally identified 1,940 genes showing differential expression between young and old animals in heads. Comparing these lists to similar lists of age-related genes from five other publications (see **Supplementary Data S4**) we found that the majority of our expression candidates had been previously identified in one or more studies. The figure shows the number of genes we identified that were unique to our study (e.g., 43 body genes), or that were found in other studies (e.g., 11 body genes were identified in all five other studies we examined).

Supplementary Data 2.1. RILs selected for RNAseq analysis. The following 20 short-lived and long-lived RILs were selected for pooled RNAseq analysis of gene expression in bodies (thorax + abdomen) of females.

Short-lived body pool:

RIL	Median Lifespan (Hrs)	Median Lifespan (Days)
21086	826.2	34.4
21105	564.1	23.5
21135	780.3	32.5
21148	926	38.6
21188	754.8	31.4
21190	801.2	33.4
21220	928.2	38.7
21241	897.6	37.4
21247	850.8	35.4
21257	1067.6	44.5

Long-lived body pool:

RIL	Median Lifespan (Hrs)	Median Lifespan (Days)
21001	1783.4	74.3
21004	1663.6	69.3
21076	1807.3	75.3
21079	1831.3	76.3

21091	1639.6	68.3
21132	1663.6	69.3
21142	1689.1	70.4
21143	1713	71.4
21168	1713	71.4
21214	1639.6	68.3

The following six RILs were used for strain-by-strain analysis of gene expression in female heads.

RIL	Median Lifespan (Hrs)	Median Lifespan (Days)
21004	1663.6	69.3
21157	972.1	40.5
21168	1713.0	71.4
21210	1713.0	71.4
21214	1639.6	68.3
21250	1043.8	43.5

Supplementary Data 2.2. FlyBase controlled vocabulary searches. We searched the FlyBase controlled vocabulary (CV) terms (http://flybase.org/static_pages/termlink/termlink.html, accessed 6 January 2016) for the words "aging", "lifespan", "lived", and "longevity", and identified 9 terms, each associated with multiple genes. The number of independent genes across all terms is 568. The "CODE" associated with each term can be used along with **Supplementary Table S4** to determine which candidate gene is associated with each term.

CODE: 1
CV ID: GO:0007568
CV term: aging
genes: 175

CODE: 2
CV ID: FBcv:0000384
CV term: aging defective
genes: 30

CODE: 3
CV ID: GO:0007569
CV term: cell aging
genes: 2

CODE: 4
CV ID: FBcv:0000791
CV term: delayed aging

CODE: 6
CV ID: GO:0010259
CV term: multicellular organismal aging
genes: 171

CODE: 7
CV ID: FBcv:0000792
CV term: premature aging
genes: 4

CODE: 8
CV ID: FBcv:0000386
CV term: long lived
genes: 264

CODE: 9
CV ID: FBcv:0000385
CV term: short lived

genes: 2

genes: 503

CODE: 5

CV ID: GO:0008340

CV term: determination of adult lifespan

genes: 168

Supplementary Data 2.3. Gene Ontology (GO) analysis summary. Gene ontology analysis was conducted via the geneontology.org site on January 27, 2016 (Analysis Type: PANTHER Overrepresentation Test - release 20150430; Annotation Version and Release Date: GO Ontology database - Released 2015-08-06). The database contained 13,690 *D. melanogaster* genes, and 245/252 (1,898/1,940) of the genes showing differential expression in bodies (heads) were present. Selected "biological process" GO category results are shown below. The "Genes" column is the total number of genes of that category in the database, "DE genes" is the number of those genes identified as differentially-expressed in our dataset, "FE" is the fold enrichment (the number of genes identified relative to the number expected by chance), and "*P*-value" is the result of a Bonferroni-corrected significance test for enrichment.

Bodies

GO Category	**Genes**	**DE Genes**	**FE**	***P*-value**
antibacterial humoral response (GO:0019731)	29	10	>5	5.49E-07
All 10 increase expression with age				
defense response to Gram-positive bacterium (GO:0050830)	41	11	>5	9.27E-07
All 11 increase expression with age				
defense response (GO:0006952)	330	28	4.74	4.25E-08
All 28 increase expression with age				
egg coat formation (GO:0035803)	14	5	>5	1.71E-02
All 5 decrease expression with age				

skeletal myofibril assembly (GO:0014866)	7	5	>5	5.93E-04
All 5 decrease expression with age				
Myofibril assembly (GO:0030239)	40	10	>5	1.15E-05
All 10 decrease expression with age				

Heads

GO Category	**Genes**	**DE Genes**	**FE**	***P*-value**
antibacterial humoral response (GO:0019731)	29	17	4.23	3.09E-03
16/17 increase expression with age				
defense response to Gram-positive bacterium (GO:0050830)	41	24	4.22	2.24E-05
All 24 increase expression with age				
defense response (GO:0006952)	330	84	1.84	4.63E-04
70/84 increase expression with age				
electron transport chain (GO:0022900)	86	42	3.52	2.21E-08
39/42 decrease expression with age				
ATP metabolic process (GO:0046034)	133	57	3.09	8.93E-10
52/57 decrease expression with age				

Supplementary Data 2.4. Extracting genes showing age-related changes in expression from previous studies.

Pletcher et al. (2002) - PMID: 12007414

1,312 genes with age-related changes in expression.

We received information on the probe sets that showed age-related changes in gene expression from the corresponding author of this paper. The results were from an updated analysis, and a slightly larger set of significant genes were identified than was reported in the original paper. Using the annotation file provided to accompany the expression array used in the study we converted probe sets to FlyBase gene IDs ("FBgn"), ignoring any probe sets lacking FBgn. FBgn were then updated to current FBgn using FlyBase (http://flybase.org/static_pages/downloads/IDConv.html, accessed 30 December 2015). Subsequently, we retained only those unique genes with a one-to-one relationship between current and former FBgn (differences among annotation releases cause a small minority of putative open reading frames to be merged or split as annotations improve).

Landis et al. (2004) - PMID: 15136717

854 genes with age-related changes in expression.

We extracted FBgn for probe sets showing increased or decreased expression with age from Table 1 (tabs "old up", "O2 OLd UP", "old down", and "old o2 down"), and used a similar protocol as described above to filter FBgn.

Lai et al. (2007) - PMID: 17196240

2,118 genes with age-related changes in expression.

We extracted FBgn for probe sets showing increased or decreased expression with age from Supplementary Table 3. Probe sets reported to be associated with more than one gene symbol and/or FBgn were ignored. Subsequently we used the same protocol as described above to update and filter FBgn.

Zhan et al. (2007) - PMID: 17623811

3,181 genes with age-related changes in expression.

We extracted FBgn for genes with age-related expression changes from Supplementary Tables 1-7, ignoring any gene lacking an FBgn number. After updating FBgn numbers via FlyBase, we merged data from all tissues. Thus, the set of genes we employ show a significant age-related change in at least one, and perhaps as many as seven tissues.

Carlson et al. (2015) - PMID: 26090231

1,527 genes with age-related changes in expression.

We extracted FBgn for genes showing changing expression with age from Supplementary Table 1, and used the same protocol as described above to filter FBgn.

Chapter III

Functional Validation of Causative Loci Contributing to Nicotine Resistance in 1st Instar Larvae

Abstract

All species possess detoxification pathways with considerable natural variation enabling plastic responses to environmental toxins. Given the widespread use of pesticides, and the frequent development of resistance to these compounds by crop pests, it is critical to understand the genetic basis of xenobiotic resistance in insects. Here, we follow-up on a previous multi-parent quantitative Trait Loci QTL (QTL) mapping study carried out in the *Drosophila* Synthetic Population Resource (DSPR) by performing quantitative complementation tests, RNAi, over-expression experiments, and CRISPR to functionally validate the effects of *Cyp281*, *Cyp28d2*, and *Ugt86Dd* on nicotine resistance and attempt to identify causative sequence variants. To determine whether these loci harbor segregating variation influencing resistance, we crossed susceptible and resistant strains to a range of deficiencies and insertional mutants. These quantitative complementation tests (QCTs) revealed functional allelic variation at *Cyp28d1* and *Cyp28d2*, and suggested that multiple UDP-glucuronosyltransferase (Ugt) genes likely harbor allelic variation with effects on nicotine resistance. Ubiquitous RNAi knockdown of *Cyp28d1* and *Ugt86Dd*, as well as targeted knockdown in the anterior midgut, reduced nicotine resistance, providing evidence these genes are involved in resistance. Sequencing the *Ugt86Dd* open reading frame revealed a 22bp coding deletion segregating in our multi-parental mapping panel, with the four most susceptible founder haplotypes all harboring the deletion allele. We constructed overexpression genotypes using both *Ugt86Dd* alleles, and found that nicotine resistance is significantly greater following anterior midgut overexpression of the insertion allele compared to the deletion allele, implying the variant has a functional role. Successful generation of custom *Ugt86Dd* mutants using the CRISPR-Cas9 system revealed *Ugt86Dd* has a large effect on nicotine resistance within the DSPR. Out of the three candidate genes functionally tested,

Ugt86Dd appears to be a causative gene and the complex 22bp deletion is likely a causative variant with variable effects on nicotine resistance in the DSPR.

Introduction

Toxins are a common and diverse biomolecule that are employed to protect plants and animals. For instance, >10% of terrestrial plant species produce some form of toxins for protection [1,2], while several animals produce toxins as a defense mechanism [3]. Thus, with the common occurrence of toxins in nature, complex mechanisms to avoid and catabolize these toxic substances have been developed by animals, such as insects and humans. These sophisticated detoxification systems allow for variable responses to endogenous and exogenous environmental toxins. Fortunately, with the recent increased employment of insecticides/pesticides against crop pest, and the considerable evolutionary pressure to frequently develop resistance to these insecticides/pesticides [4] allows for a unique opportunity to better understand how detoxification works on a genetic level. Thus, it is critical to understand the genetic basis [5] of xenobiotic resistance in insects to help design valuable crop pest management strategies.

Broadly, we know the biochemical pathway that catabolizes and excretes xenobiotics from the cell [6]. This biochemical pathway can be broken down into three phases. Phase I includes detoxification enzymes known as cytochrome P450 genes (P450s) that carry out oxidation reactions on a broad range of xenobiotics and catabolize these xenobiotics into various levels of toxic by-products. To date, *Drosophila melanogaster* has roughly 100 P450 genes [7], of which are mainly expressed in the fatbody midgut/hindgut, and malpighian tubules [7,8]. Next, large functional groups are added to the by-products by phase II enzymes such as glutathione-S-transferases (GSTs) and UGTs to improve the molecules hydrophilic properties for excretion [9]. Lastly, phase III includes transporters, such as ATP-binding cassette (ABC) transporters, that remove bulky hydrophilic by-products from the cell. Even with this knowledge and identification of many detoxification genes [7,10], our understanding of the precise pathway

of detoxification for each compound and the variable genetic affects in xenobiotic metabolism are largely unknown.

To tease apart the genetic contributions that are associated with xenobiotic resistance, traditional mapping approaches have been typically employed, such as Genome-wide association studies (GWAS) and QTL mapping, with limited success in elucidating causative genes. Typically, GWAS compare cases and controls for a trait, and have had some success in elucidating common genetic variants associated with drug resistant [11–14]. However, a majority of GWAS are small in size and are severely underpowered to detect true genetic variants, including rare disease causing alleles [15–17]. Traditional QTL mapping takes parents that segregate differentially for the trait under test, and phenotypes progeny that are subjected to very few rounds of meiotic recombination. This enables QTL mapping to achieve strong statistical power to detect segregating regions, but limits mapping resolution to several Centimorgans that can harbor thousands of genes [18–21]. To overcome the weaknesses of both approaches, a series of multi-parental mapping panels were developed in a variety of model organisms to facilitate the ability to find causative genes. These multi-parental mapping panels include the *Arabidopsis thaliana* MAGIC lines [22], the maize Nested Association Mapping population (NAM) panel [23], the mouse collaborative cross [24], the Diversity Outbred mouse population [25], and the DSPR [26]. These multi-parental mapping panels will allow scientists to dissect the variable genetic effects of xenobiotic resistance, such as nicotine resistance.

Initially, phenotypic variation in nicotine lethality was demonstrated in in genetically different *Drosophila* strains [27]. Moreover, studies performing various screens on wild-derived lines have uncovered genetic variation that mediates phenotypic variation for nicotine resistance [28–30]. Recently, the DSPR [26,31] was employed to dissect genetic variation that affects nicotine resistance [32]. Phenotyping a majority of the DSPR (1274 RILs), four mapped QTL

were uncovered that explained 70% of the broad sense heritability. Interestingly, QTL1 explained roughly 9% of the broad sense heritability and harbored two phase I enzymes *Cyp28d1*, and *Cyp28d2*. Moreover, QTL4 roughly explained 50% of the broad sense heritability and harbored ten phase II UGT genes. A genome-wide RNA-seq experiment demonstrated both P450 genes under QTL1, *Cyp28d1* and *Cyp28d2*, and one of ten UGT genes under QTL4, *Ugt86Dd*, were differentially-expressed between susceptible and resistant DSPR founders. These candidate genes are ripe for validation.

Here, we employed an array of functional assays that include QCTs, RNAi, over-expression, and CRISPR to functionally validate candidate genes uncovered in a nicotine resistance screen by Marriage *et al.*, (2014). QCTs revealed functional allelic variation at *Cyp28d1* and *Cyp28d2*, and suggested that multiple nicotine resistance factors are present within the QTL4 region, consistent with the idea that several of the Ugt genes are involved. We found ubiquitous RNAi knockdown of *Cyp28d1* and *Ugt86Dd*, as well as targeted knockdown in the anterior midgut, reduced nicotine resistance, providing evidence these genes are involved in nicotine resistance. While homing in on *Ugt86Dd*, a complex 22 basepair InDel was discovered. Over-expression experiments of the insertion allele and deletion allele found that nicotine resistance is significantly greater following anterior midgut overexpression of the insertion allele compared to the deletion allele, implying the variant has a functional role. Lastly, *Ugt86Dd* CRISPR mutants on average were more susceptible than nonCRISPR mutants.

Materials and Methods

Drosophila Stocks and Maintenance:

All flies utilized for experiments were raised and maintained on standard cornmeal–yeast–molasses medium at 25 °C on a 12:12 hour light/dark cycle with constant humidity of 50% unless otherwise stated. Briefly, 4-5 males and 10 virgin females and were harvested under CO_2 anesthesia in regular narrow, polystyrene fly vials and allowed to recover for at least a day before initiating crosses. Stocks employed for QCTs were selected from DrosDel, Exelixis, BSC, and gene disruption project collections [33–36] housed at the Bloomington Drosophila Stock Center, Bloomington, Indiana (Table 3.1). These collections are a creation of Deficiency (*Df*) stocks with characterized molecular breakpoints and mutations in positional candidate genes that have been generated in the w^{1118} isogenic background strain (Table 3.1). Stocks utilized for RNA interference (RNAi) studies were selected from the KK, GD [37] and Transgenic RNAi Project (TRiP) RNAi libraries housed at the Vienna Drosophila Resource Center (VDRC) or Bloomington Drosophila Stock Center, Indiana (Table 3.1). Quickly, the VDRC collections were created by randomly inserting an hpRNA into the genome via p-elements or by targeting the hpRNA to specific landing sites via ΦC31, while the TRiP collection targets shRNA to specific landing sites via ΦC31. Furthermore, GAL4 drivers utilized to turn on the UAS-RNAi stocks were purchased from the Bloomington Drosophila Stock Center, Bloomington, Indiana or were donated by flygut [38] (Table 3.1). Lastly, stocks utilized for CRISPR experiments where generated from taking BestGene's Vas-Cas9 crossing in the 3rd chromosome of A4 from the DSPR (Table 3.1).

Larval Nicotine Resistance Assay in Polystyrene Fly Vials:

This larval nicotine resistance assay utilized in this study was adapted from Marriage *et al.*, (2014) larval nicotine resistance assay. All parts of the assay are similar expect larvae collections. Following egg laying, 30 1st instar larvae were collected across a series of parent vails per each cross, unless otherwise stated. Each set of 30 1st instar larvae were place on either standard fly media or standard fly media supplemented with 0.18 ul/ml nicotine (N3876, Sigma) made within the last 18 hours in regular narrow polystyrene fly vials (for preparation of food see Marriage et al., 2014). Eleven-to-thirteen days later the number of adults that had emerged in each vial were counted via CO_2 anesthesia. By this time all pupae on the sides of the vial had hatched.

Dominance Assay:

To estimate the effect size of each chromosome has on nicotine resistance, A3 and A4 third chromosomes where swapped between their respective backgrounds using a series of balancers. The final genotypes were listed in the order of X/Y;2;3 and consisted of A4/A3;A4;A3 or A3/A4;A3;A4. For a detailed crossing scheme to create these genotypes see Supplementary Data 4. To determine the dominance state of the 3rd chromosome, these newly created stocks were backcrossed to either A3, A4, or self-crossed and compared to self-crossed A3 and A4 for nicotine resistance using the larval nicotine resistance assay in polystyrene fly vials. A total of 2 biological replicates per cross were recorded on control food, while 6 biological replicates per cross were documented on nicotine-laced food.

Quantitative Complementation Test:

To narrow down the genomic loci and functionally test genes implicated in our mapping and RNAseq experiments, we employed QCTs. In order to perform QCT on 1st instar larva, we

replaced the original balancer with Cyo^{GFP} or $TM6^{GFP}$. We crossed males that either contained the insert or defiency to A4 (resistant line) or A3 (susceptible line) females. Flies were allowed to mate for 48 hours in regular narrow, polystyrene fly vials. From here, a larval nicotine resistance assay in polystyrene fly vials was performed utilizing 9-34 1st instar larvae per replicate on either control or nicotine-laced food. A total of 2-8 biological replicates per cross were recorded on control food, while 8-18 biological replicates per cross were documented on nicotine-laced food. The average viability per cross was recorded, and analyzed/interpreted using the reduced model: $y = \mu + L + G + (L \times G) + R(L \times G)$ [39,40,40].

RNAi Assay:

To functionally examine the effects of nicotine resistance on *Ugt86Dd*, *Cyp28d1*, and *Cyp28d2* uncovered in marriage *et al.*, (2014) RNAseq data, we employed the binary GAL4-UAS RNAi system. We utilized VDRC p-element (GD), ΦC31 (KK), and the ΦC31 TRIP integrated UAS transgenes along with their respective controls (Supplemental Table 1). We examined phenotypic effects associated with ubiquitous or tissue specific RNAi knockdown of selected genes. We took males that ubiquitously or tissue specifically express GAL4 through their respective promoters and crossed to females of each control and each UAS-RNAi strains. Appropriate F_1 1st instar larvae were selected and utilized to perform the larval nicotine resistance assay in polystyrene fly vials. A total of 3-5 biological replicates per cross were recorded on control food, while 8-10 biological replicates per cross were documented on nicotine-laced food.

***Ugt86Dd* DSPR InDel PCR Detection Assay:**

Primer set 5'-CGCTTTTGCTCAGCATTTTA-3' and 5'- ATATGTGGCAGGTGAACGAA-3' was employed to amplify the InDel variant (219 or 197 base pair product) at the *Ugt86Dd* locus

in the DSPR under cycling parameters: 95°C 2 mins, 35 cycles of 95°C 20s, 57°C 25s, 72°C 30s, and a final 2 min extension at 72°C. To properly call the lnDel variant, a 2% agarose gel was utilized and ran in 1x TAE at 50v for 90 minutes, changed to new 1xTAE buffer and ran for 60 minutes at 100v. To verify InDel variation calls in the DSPR, DNA PCR products were cleaned up using and following manufacture protocols for a QIAquick Gel extraction Kit (Cat #: 28704; Lot: 436176125) or a QIAquick PCR Purification Kit (Cat #: 28104; Lot: 148024523), and sent to ACGT for sequencing and aligned utilizing DNAstar packages.

***Ugt86Dd* Over-Expression Assay:**

To complement our RNAi validation assays, a gain-of-function analysis was performed on *Ugt86Dd*. PCR was performed using template DNA from A3 and A4 of the DSPR. To insert AttB sites flanking *Ugt86Dd* from the selected DSPR founders, primer set 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTTACAACATGAGATTATTAACTGTGATCGCGA-3' and 5'-GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATGTTTCTTAAGCTTATCAG-3' were utilized. Following the manufacture protocol, the PCR Cloning System with Gateway® Technology from Invitrogen (Cat #: 12535-029) was used to create an entry vector (pDONR221) via the BP reaction. All selected entry clones were Sanger sequenced on both ends using primer set M13F(-20) (5′-GTAAAACGACGGCCAGT-3′) and M13R (5′-GGAAACAGCTATGACCATG-3′) to verify length and direction of gene. The destination vector pUASg.attB was generously donated by FlyORF [41] and used in combination with the LR reaction (Cat #: 11791-020) to generate pUASg.Ugt86Dd_A3.attB and/or pUASg.Ugt86Dd_A4.attB expression clones. All selected expression clones were Sanger sequenced using primer hsp-GW-F (5'-GCAACTACTGAAATCTGCCAAG-3') to verify

direction of insert. To create over-expression *Drosophila* stocks containing the various expression constructs, BestGene injected the stock M{3xP3-RFP.attP}ZH-86F(with{vas.int.DM}ZH-2A) with the created expression plasmids at concentrations of 0.510 – 0.515 ug/ul. Once we received over-expression stocks, the X and 2nd chromosomes segregated for different chromosomes, while the 3rd chromosome containing the injected construct was homozygous for the third chromosome. These stocks were utilized in conjunction with tissue specific drivers to perform larval nicotine resistance assay in polystyrene fly vials. A total of 2-3 biological replicates per cross were recorded on control food, while 6 biological replicates per cross were documented on nicotine-laced food.

***Ugt86Dd* DGRP InDel Population Association Assay:**

In the course of sequencing the DSPR founders, we observed a SNP (G/C) segregated perfectly with the *Ugt86Dd* InDel. We wanted to determine if the Drosophila Genetic Reference Panel (DGRP) harbored this InDel variant. Thus, using the DGPR2 (http://dgrp2.gnets.ncsu.edu/) the observed (G/C) SNP was used as a proxy to identify DGRP strains that differentially segregate for the InDel variant in *Ugt86Dd* [42,43]. Once lines were selected, the *Ugt86Dd* DSPR InDel PCR detection assay was used to confirm seven homozygous DGRP strains (28199, 25177, 28185, 28213, 25206, 25176, 37525) contained the *Ugt86Dd* deletion and 7 random homozygous DGRP strains (28239, 25200, 28160, 28226, 28197, 25174, 28295) contained the *Ugt86Dd* insertion. To create populations fixed for the InDel variant, all possible inter-line crosses were performed on DGRP strains homozygous for the deletion variant and DGRP strains homozygous for the insertion variant using 10 virgin females and 10 males from each strain. Ten F_1 virgin females and males from each corresponding cross were selected and utilized to initiate their respective populations. Each population was maintained in 1 Gallon glass bottles at room

temperature in consist light. Each population was tipped to new food every 10-14 days. To evaluate nicotine resistance on these populations, a larval nicotine resistance assay in polystyrene fly vials was performed on 600 F7 larvae from each population.

CRISPR/Cas9 Experiment:

To optimize our CRISPR guideRNA (gRNA), http://tools.flycrispr.molbio.wisc.edu/targetFinder/ [44] and http://crispr.mit.edu/ [45] was employed. The target sequence TCACTACGAAGTCATTGTGGAGG was a top hit within both software's and is close to the InDel variant of interest. Following the U6-gRNA (chiRNA) cloning protocol from flyCRISPR [46], cloning *Ugt86Dd*_sense and antisense primers (5'Phos-CTTCGTCACTACGAAGTCATTGTGG-3' and 5'Phos-AAACCCACAATGACTTCGTAGTGAC-3') was utilized to perform the cloning method laid out in the U6-gRNA (chiRNA) cloning protocol. To verify transformation of the U6-gRNA, sequenced plasmid purified via a QIAprep spin Miniprep Kit (Cat #: 27104; Lot: 151010133) following manufacture protocol was sent to ACGT. Plasmid (0.25ug/ul) containing our gRNA was sent to BestGene for injection into y[1] M{vas-Cas9}ZH2A w[1118]; ; A4. To verify CRISPR mutants, primer set 5'-ACGCTTTTGCTCAGCATTTT-3' and 5'-GGCTGGGGATACCATTTCTT-3' was utilized to amplify a region containing our CRISPR target site under cycling parameters: 95°C 2 mins, 35 cycles of 95°C 20s, 57°C 25s, 72°C 30s, and a final 2 min extension at 72°C. For each reaction 10ul PCR product, 7ul molecular grade water, and 2ul NEB buffer 2 (Cat #: B7002S; Lot: 0181508) were incubated in a heat block set at 95°C for 5 mins, and allowed to cool to room temperature for ~2 hours. Then 1ul of T7 endonuclease (Cat #: M0302L; Lot: 0041512) per reaction was added and incubated at 37°C for 15 mins. After 15 mins, 2ul per reaction of 0.25mM EDTA was added to stop the T7

endonuclease activity, and immediately loaded all the reaction into a 1.5% agarose gel to perform gel electrophoresis in 1% TAE to confirm positive CRISPR mutants.

We discovered the injection strain y[1] M{vas-Cas9}ZH2A w[1118] strain was not homozygous at the *Ugt86Dd* loci, thus a custom injection strain was created (y[1] M{vas-Cas9}ZH2A w[1118]; ; A4). Using this custom strain, plasmid pBbsI (0.25ug/ul) containing our gRNA (TCACTACGAAGTCATTGTGGAGG) was sent to BestGene for injection. Of the 300 injected eggs ~ 100 larvae survived the injection. From these larvae, a total of 53 G0 males and 36 G0 females were harvested. Of the 99 crosses set up, 57 (32 female G0 and 25 male G0) crosses produced viable offspring. Fifty percent of the G0 animals tested (18/36) via T7 endonuclease led to at least 1 putatively CRISPR edited F_1 animal. Per G0, the fraction of F_1 animals that are positive was ~22%, but if we consider just the 18 putatively positive CRISPR mutants containing a least 1 edited F_1, the percentage of F_1 animals that are positive was ~44%. Sequencing was employed to verify our T7 endoclease assay results, and a total of 16 unique CRISPR induced mutations were created with a 2bp deletion the most common (Supplementary Figure S8). Furthermore, all CRISPR induced mutants demonstrated a premature stop codon no more than 120bp after the CRISPR event, except for two in frame mutants (3bp & 6bp) and the complex 6bp insert (Supplementary Data 3.3). In total we developed 16 CRISPR stocks with unique CRISPR events and 7 CRISPR stocks that went through the CRISPR process but yield no mutation. These CRISPR stocks are homozygous for the X and 3rd chromosome, but segregate for the 2nd chromosome. These stocks were utilized for larval nicotine resistance assays in polystyrene fly vials to compare CRISPR mutants to CRISPR non-mutants. A total of 5 and 6 biological replicates per cross were recorded on control or nicotine-laced food.

Statistical Analysis:

All statistics used in this study were performed using the R statistical programming language (www.R-project.org) via base script packages or custom written scripts.

Ugt86Dd InDel Meta-Population Screen:

To verify the *Ugt86Dd* InDel in natural populations, we retrieved raw sequences of 14 pooled population samples from NCBI SRA (BioProject accession PRJNA256231, [47]), and see Supplement Table S11 for accession numbers of individual libraries). For each population, we counted a relative frequency of a particular insertion sequence and two particular deletions sequences. For the insertion sequence, we used the 20bp sequence centered on the insertion 'ATTGTGGAGGACATTCATCG'. While the first deletion sequence includes the 20bp sequence centered on the deletion 'ACTACGAAGTGAATTCGTTC'. The second deletion sequence harbors a SNP at position 632 and is found in half of the deletion sequence 'ACTACGAAGTGATTTCGTTC'.

Results

Most of the phenotypic signal for nicotine resistance is on the 3rd chromosome

Our previous data suggest QTL4 on chromosome 3 explains ~45-50% of the broad sense heritability [32]. Thus, the 3rd chromosome in theory should have an immense effect on nicotine resistant within the DSPR. To test the genotypic effect the 2nd and 3rd chromosome have on nicotine resistance, stocks were constructed that consisted of A4/A3;A4;A3 or A3/A4;A3;A4 (see materials and methods for creation). These newly created stocks were backcrossed to A3, A4, or self-crossed and compared to self-crossed A3 and A4 for nicotine resistance. It was observed most of the genotypic signal for nicotine resistance appears to be harbored on the 3rd chromosome and having one functional copy of the A4 third chromosome within A3 recovers to A4 like resistance, indicating the genotypic contribution appears to be non-additive but dominant (Supplementary Figure 3.S1, Supplementary Figure 3.S2). However, this dominant phenotype could be caused by multiplicative interactions between unlinked rare variants that are additive on the 3rd chromosome or simply multiplicative interactions in strong LD under QTL4 that are additive and will be of great importance to establish such mutations increase risk in a linear or synergistic manner [48]. Lastly, a Welch's Two Sample *t*-test was performed between self-crossed A3 and A4-DELxA4-DEL (t= -2.035, p= 0.069) or A4-DELxA3 (t= -2.068, p= 0.066), both produced a marginally nonsignificant result indicating a trend. This trend suggest the effects between A3 and A4 on the 2nd chromosome may be additive and have a small effect on nicotine resistance between A3 and A4. Taken together this data suggest there is a series of variants that may or may not be in LD on the third chromosome that contribute greatly to nicotine resistance, while the second chromosome has a small effect on the phenotype within the DSPR.

Deficiency and mutation complementation testing

The theory on QCTs is straightforward, take deficiencies or mutant alleles of a candidate genes and test their associations with naturally derived alleles [49,50]. If the candidate region/gene contains natural allelic variation for the phenotype, natural alleles may fail to complement. Thus, using this ideology we employed QCTs on available deficiency and mutations in selected regions and positional candidate genes (*Cyp28d1*, *Cyp28d2*, *Ugt86Dd*, *Ugt86Dh*, *Ugt86Dj*) under our previously mapped QTL1 and QTL4 [32]. To assess failure to complement, nicotine resistant assays were performed on the four selected genotypes: mutant or deficiency/A3, mutant or deficiency/A4, background/A3, and background/A4.

An ANOVA was used to assess complementation for deficiencies and positional mutants employed under QTL1. All demonstrated a significant interaction indicating failure to complement (Figure 3.1a). Df(2L) Exel6011 ($F_{1,28}$= 62.3, p< 0.005), Df(2L) BSC693 ($F_{1,28}$= 121.8, p< 0.005), Mi{ET1} Cyp28d2^{MB02776} ($F_{1,52}$= 77.1, p< 0.005), Cyp28d1^{MB03293} ($F_{1,52}$= 11.5, p< 0.01). These data suggest that *Cyp28d1* and *Cyp28d2* are likely candidate Quantitative Trait Gene (QTGs) harboring variation affecting nicotine resistance between A3 and A4.

Using the same model above, deficiencies employed under QTL4 all demonstrated a significant interaction and failed to complement; however, positional mutants demonstrated complementation (Figure 3.1b). Df(3R) ED5506 ($F_{1,31}$= 78.4, p< 0.005), Df(3R) Excel7806 ($F_{1,36}$= 40.8, p< 0.005), Df(3R) Excel8152 ($F_{1,35}$= 22.2, p< 0.005), Mi{ET1} Ugt86Dj^{MB04890} ($F_{1,34}$= 4.0, p= 0.052), Ugt86Dh^{MB11311} ($F_{1,35}$= 0.9, p= 0.34). These data suggest that deficiencies Df(3R) ED5506 and Df(3R) Excel7806 uncovering the region harboring the previously implicated candidate gene *Ugt86Dd*, contains at least one QTG that harbors allelic variation for nicotine resistance between A3 and A4. The deficiency Df(3R) Excel8152 uncovering the genomic region containing the other nine UGT genes, likely harbors one segregating allelic

variant for nicotine resistance between A3 and A4 . Furthermore, our data suggest out of the 10 UGTs mapped under QTL4, *Ugt86Dj* and *Ugt86Dh* are most likely not QTGs that harbor variation affecting nicotine resistance between A3 and A4. However, *Ugt86Dj* was marginally nonsignificant and with a higher sample size might reveal *Ugt86Dj* as a minor segregating variant between A3 and A4. Lastly, we can not rule out the possible the inserts do not abolish the function of the gene within the insertional mutants. However, this is likely not the case for *Ugt86Dj* or *Ugt86Dh*, both inserts are harbored in the first exon of each gene.

Ubiquitous depletion of *Ugt86Dd*, *Cyp28d1*, and *Cyp28d2* demonstrates a role in nicotine resistance

Four candidate genes (*Cyp28d1*, *Cyp28d2*, *Ugt86Dd*, and *Ugt86Dj*) under QTL1 and QTL4 [32] were selected for GAL4-UAS RNAi depletion (knock-down) experiments. Initially, we took males that have ubiquitous expression of GAL4 over a balancer chromosome and crossed to virgin females that are isogenic for the targeted UAS RNAi constructs, and performed nicotine resistance assays on F_1 larvae. When assessing viability, two genotypes emerged Balancer/UAS RNAi and GAL4/UAS RNAi. We took the fraction of GAL4/ UAS RNAi that emerged from each depletion strain and compared to the control strain. Using this ideology, all GAL4-UAS strains except the GAL4-UAS targeting *Ugt86Dj* showed a significant reduction in nicotine resistance compared to the control (Supplementary Figure 3.S3). We could not rule out the chance that the difference was due to specific genotypes in the various strains emerging at different rates. To deal with this potential problem, we replaced all balancer chromosomes with a balancer chromosome marked with dominant GFP or YFP and selected F_1 1^{st} instar larvae of the correct genotype. If we assess the possibility genetic manipulations performed on our RNAi genotypes reduce survival, we examined normal survival of the RNAi genotypes on regular food.

A Welch's *t*-test for each genotypes demonstrates no abnormalities appear to cause no significant reduction in survival (Figure 3.2a). When performing RNAi nicotine resistance assays, we see all GAL4-UAS RNAi strains demonstrate a significant reduction in nicotine resistance in comparison to the controls (Figure 3.2b). Thus, concluding the reduction in nicotine resistance that is observed, small, is likely a true biological signal. Furthermore, we hypothesize that if our selected targets are important for nicotine resistance, we should be able to employ various RNAi backgrounds and see similar phenotypes. Interestingly, we were able to replicate this significant reduction in nicotine phenotype across various backgrounds (GD, KK, TRiP) further suggesting that *Cyp28d1*, *Cyp28d2*, and *Ugt86Dd* likely play a role a functional role in nicotine resistance (Figure 1).

RNAi lines targeting the Anterior Midgut demonstrate role in nicotine resistance

The digestive tissue has been characterized in *Drosophila* and shown to have variable effects on xenobiotic metabolism and may impact nicotine resistance [30,51–57]. The digestive tissue consists of the midgut, hindgut, and malpighian tubules. Of interest is the midgut that can be divided into three compartments that are composed of several subsections [56,58] of which metabolism of various xenobiotics may occur through the length of the midgut via a complex interaction of subsections [56]. The midgut is a likely spot to encounter nicotine-laced food after ingestion and our candidate genes are highly expressed in the anterior larvae midgut [56]. Using males that were homozygous for the tissue specific GAL4 and virgin females that were homozygous for the UAS RNAi constructs, every F_1 1st instar larvae selected targeted RNAi to deplete *Ugt86Dd*, *Cyp28d1* and *Cyp28d2* within specific subsets of tissues: fatbody, midgut, hindgut, and malpighian tubules in the VDRC GD background. It was observed that depletion of the selected genes in the fatbody has no effect on resistance (Supplementary Figure 3.S4).

However, tissue specific depletion of *Cyp28d1* and *Ugt86Dd* within the anterior and posterior midgut revealed a significant reduction in nicotine resistance in the anterior midgut but not the posterior midgut (Figure 3.3a). To confirm our GD library results, tissue specific depletion of was performed in the VDRC KK and TRiP backgrounds. Tissue specific depletion within the anterior and posterior midgut in KK and TRiP backgrounds confirmed *Cyp28d1* and *Ugt86Dd* likely have a role in the anterior midgut (Figure 3.3b, Figure 3.3c). If we assess the possibility a reduction in nicotine resistance is due to minor abnormalities in our RNAi genotypes on regular food, there appears to be no abnormalities that cause a significant reduction in nicotine resistance. However depletion assays in the malpighian tubules with multiple drivers demonstrated an increase in resistance across several drivers (Supplementary Figure 3.S4). This is opposite in what our previous data has shown and may indicate a complex feedback mechanism. Taken together, our data suggest *Ugt86Dd* and *Cyp28d1* likely influence nicotine resistance within the anterior midgut independent of genetic background, and *Ugt86Dd* and *Cyp28d1* appear not to function in the posterior midgut.

DSPR *Ugt86Dd* InDel Variant Correlative for nicotine resistance

When attempting to construct an over-expression system involving *Ugt86Dd*, it was discovered that *Ugt86Dd* has a 22bp deletion allele harbored inside the second exon of the coding region in A3 that causes a frameshift mutation that results in a nonfunctional protein. To check the frequency of the InDel variant in the DSPR, PCR was employed. PCR analysis suggested DSPR founders A3, AB8, B6, and B7 all contained an isogenic copy of the deletion allele, while all other DSPR founders contained an isogenic copy of the insert allele (Supplementary Figure 3.S6). To confirm our PCR analysis, we Sanger sequenced the PCR products for each of the DSPR founders. Sanger sequencing confirmed the predicted results of the PCR analysis.

Furthermore, looking at the haplotype means for QTL4 from marriage *et al.*, (2014), we can observe that all nicotine susceptible founders harbored the isogenic deletion allele (Supplementary Figure 3.S7). We conclude that the *Ugt86Dd* InDel variation is common in the DSPR and correlates with nicotine resistance.

Over-expression of InDel variant increases nicotine resistance

To nicely complement our knock-down experiments and corroborate the InDel variant is plausibly important, gain-of-function analyses were performed using the constructed over-expression genotypes for both *Ugt86Dd* InDel alleles. The major difference between the sets of lines, were one contained a 22bp deletion the likely is a functional null, while the other lines harbored the wildtype allele with normal expression. These lines were crossed to males expressing ubiquitous GAL4. Performing nicotine resistance assays on F_1 1st instar larvae, it was observed that all over-expression lines containing the insert construct died by 3rd instar larvae (observation), while the line containing the deletion allele survived fine to adulthood. This suggest two things, 1) ubiquitous over-expressed *Ugt86Dd* inset allele is poisonous to the organism, and 2) the deletion allele in *Ugt86Dd* is most likely a functional null allele. To deal with this problem, tissue specific over-expression was performed under the assumption that the insert allele in *Ugt86Dd* should provide a higher resistance when compared to the deletion allele. Thus, tissue specific over-expression of *Ugt86Dd* using multiple drivers within the anterior midgut demonstrated a statistically significant increase in nicotine resistance in all inset allele-containing lines (Figure 3.4a, Figure 3.4b,). Moreover, tissue specific over-expression of *Ugt86Dd* within the posterior midgut increased nicotine resistance in all inset allele-containing lines (Figure 3.4c). However, this data should be taken with caution. *Ugt86Dd* is low-to-moderately expressed in the posterior midgut [56] and likely does not have a natural role in this

tissue, but the data suggest *Ugt86Dd* is functionally important. Lastly, over-expression experiments conducted within the malpighian tubules increased nicotine resistance as similarly seen in our depletion experiments (Supplementary Figure 3.S10), suggesting a complex feedback mechanism may exist for xenobiotic resistance. Taken together, our data suggest *Ugt86Dd* likely influences nicotine resistance within the anterior midgut, posterior midgut, and malpighian tubules. Furthermore, the InDel variant appears to be functionally important for nicotine resistance.

DGRP *Ugt86Dd* InDel variant correlative for nicotine resistance

To test the idea that the InDel in *Ugt86Dd* is important for nicotine resistance, we employed strains from the DGRP [42,43]. We took the 7 strains containing the deletion variant and another 7 random strains that contained the insert variant (Supplementary Data 3.1) were challenged to nicotine and assessed the association the InDel variant has on nicotine resistance with an ANOVA ($F_{1,54}$= 1.1, p= 0.31) in the DGRP (Supplementary Figure 3.S13). This data suggested no association was present; however, a trend toward susceptibility was seen if the strain is homozygous for the deletion allele. A nonsignificant result could be due to the small sample size employed. To potentially deal with this problem, populations were created and fixed for the known InDel variant and allowed to outbred for 7 generations (Supplementary Data 3.1) in milk bottles. Six hundred F_7 1st instar larvae were measured from each population for nicotine resistance and found on average the nicotine resistance mean of the deletion population was 0.8933 percent, while the nicotine resistance mean of the insert population was 0.9350 percent (Supplementary Table S10). Using a one-way ANOVA to assess the data, we found a significant result ($F_{1,38}$= 7.1, p= 0.011) (Supplementary Figure 3.S14). We can not rule out the chance another site in the genome is also fixed in the populations, since only 7 strains were used each

create their respective population. Thus, we conclude that having the deletion allele is correlated with increase nicotine susceptibility, and the InDel variant segregates at a low frequency in the DGRP in comparison to the DSPR.

CRIPSR confirms *Ugt86Dd* and variant's role in nicotine resistance

We aimed to delete a portion of the *Drosophila Ugt86Dd* InDel site by CRISPR–Cas9 mediated DNA cleavage to phenocopy the A3 nicotine resistance phenotype. This would give compelling evidence that *Ugt86Dd* is functionally active and the InDel variant is likely a variant contributing to nicotine resistance in the DSPR. First instar Larvae from nonCRISPR mutants and CRISPR mutants (Supplementary Data 3.2)were challenged to nicotine and control food. Comparing the control versus nicotine-laced food for the individual CRISPR and nonCRISPR genotypes and Bonferroni correcting the data, 11/16 CRISPR strains provide a significant decrease in nicotine resistance (Figure 3.5a). However, looking at the data, an obvious decrease in resistance is observed for each CRISPR genotype, nicely demonstrating the over correction problems of Bonferroni. If the data is portioned into the various stop codons and tested for resistance, all stop sites demonstrate a significant decrease in nicotine resistance (Figure 3.5b). Moreover, if the raw data is portioned into CRISPR versus nonCRISPR mutants, a one way ANOVA depicts a significant decrease in nicotine resistance for all CRISPR mutants ($F_{1,92}$= 181.41, p= 2.2^-16). When comparing nonCRISPR and CRISPR mutants on normal food, no difference in larvae to adult hatching was observed ($F_{1,69}$= 1.4488, p= 0.2312) (Figure 3.5c), suggesting any induced CRISPR mutations have little effect on normal viability. More importantly, our data implies that *Ugt86Dd* likely explains a good portion of QTL4's broad sense heritability. We conclude that *Ugt86Dd* is a causative gene with variable effects for nicotine resistance in the DSPR, our data

presents strong indirect evidence that the complex 22bp InDel is likely a causative variant for nicotine resistance in the DSPR.

Discussion

Insecticides and pesticides are commonly used to ward of crop pest, such as insects. Thus, with selective pressure evident in nature to develop resistance to employed compounds [4], metabolism of specific compounds likely involves a large number of genes with variable effects. Insects possess a diversity of genes with sophisticated roles for metabolizing the plethora of xenobiotic compounds found in nature [59,60]. In this study, we employed an array of functional assays on *D. melanogaster* to identify what candidate loci under QTL1 and QTL4 identified in Marriage *et al.*, (2014) are likely naturally-occurring causative loci responsible for variable effects on nicotine resistance in the DSPR. This approach enables a systemic way to go from QTL to causative gene(s), and is the first approach to employ an array of functional assays to elucidate natural-occurring variants for nicotine resistance in *D. melanogaster*. This new information will shed insights into the variable effects genes have on xenobiotic resistance.

Here, we employed A3 and A4 that differentially segregate for nicotine resistance in the DSPR at QTL1/QTL4 [32], and found segregating variation for nicotine resistance in phase I enzymes, *Cyp28d1* and *Cyp28d2*. Unfortunately, little is known how these genes function in *Drosophila*, but DRSC Integrative Ortholog Prediction Tool (DIOPT) [61] on flybase demonstrated the most likely human homolog is *TBXAS1*. *TBXAS1* has been shown to have an effect on nicotine and smoking in humans [62,63]. To confirm this conserved function, ubiquitous RNAi on *Cyp28d1* and *Cyp28d2* were performed. Our data suggest *Cyp28d1* and *Cyp28d2* are likely functionally important for nicotine resistance. Only one study-to-date has examined natural genes/variants effects on nicotine resistance. Li *et al.*, (2012) demonstrated the Accord transposable element in *Cyp6g1* is responsible for enhanced nicotine resistance. Moreover, other xenobiotic resistance studies have uncovered transposable elements, Accord

LTR, Bari-1, Doc non-LTR, HMS-Beagle, and P, near P450s suggesting regulatory variation is critical in providing variable effects in resistance [64–70]. Neither A3 nor A4 harbor the Accord transposable element in *Cyp6g1* or any annotated transposable elements in or near *Cyp28d1* and *Cyp28d2*. However, it is possible these genes do harbor unannotated transposable elements that increase resistance. Another plausible variant are structural variants like copy number variant (CNVs), and have been shown to be associated with drug resistance [70,71]. For instance, Najarro *et al.*, (2015) has recently shown that CNV, a structural variant, at the *Cyp12d1* within the DSPR is likely a causative gene/variant for caffeine resistance. With no clear variants explaining *Cyp28d1* and *Cyp28d2* role in nicotine resistance, the availability of PacBio® sequence data in the near future will shed light on what variants are in or near *Cyp28d1* and *Cyp28d2*. Moreover, future functional assays like CRISPR will required to confirm *Cyp28d1* and *Cyp28d2* role in nicotine resistance.

Intriguingly, *Cyp28d1* and *Cyp28d2* are part of the diverse P450 superfamily that is expressed in the midgut [56]. Likely tissues natural selection acts to create variable xenobiotic detoxification effects are the midgut, the fat body, and the malpighian tubules [53,56,72–78]. Intriguingly, *Cyp28d1*, *Cyp28d2* and *Ugt86Dd* are expressed in the midgut [56]. Here, our functional RNAi assays provide compelling evidence the larvae midgut, larvae malpighian tubules, and possibly the posterior midgut likely function in xenobiotic detoxification, specifically nicotine detoxification. Upon depletion of *Cyp28d1* and *Ugt86Dd* within the anterior midgut with independent RNAi libraries and Gal4 drives, a significant decrease in resistance was observed. Furthermore, over-expression of the functional copy of *Ugt86Dd* within the anterior midgut demonstrated a significant increase in resistance. Our data for the posterior midgut is conflictive, with depletion of *Cyp28d1* and *Ugt86Dd* demonstrating no difference in phenotype, but over-expression demonstrates a significant increase in resistance. However, *Ugt86Dd* and

Cyp28d1 are lowly to moderately expressed in the posterior midgut [56] and depleting them might not have an effect on the phenotype, but over-expressing these genes might increase the ability catabolize nicotine and potentially demonstrate a phenotype. Thus, *Ugt86Dd* within the anterior midgut and posterior midgut likely has a functional role and is actively under natural selection to create variable xenobiotic detoxification effects.

Surprisingly, when *Cyp28d1*, *Cyp28d2*, and *Ugt86Dd* are individually depleted within the malpighian tubules, a significant increase in resistance was observed compared to control in multiple independent Gal4 drives. Moreover, over-expression of *Ugt86Dd* demonstrates a significant increase in resistance. This suggest a feedback loop maybe present in the malpighian tubules to handle the nicotine toxins. It is established that the malpighian tubules have a complex and diverse role in xenobiotic metabolism [79,80], and would not be far fetched for negative or positive feedback loops to exist in metabolism of xenobiotics. For instance, our lab has observed increased caffeine resistance when depleting a candidate gene in the Malpighian tubules (data not shown). Another study demonstrated depletion of *Keap1* in the malpighian tubules increases malathion resistance through a negative feedback mechanism [81]. Lastly, cancers such as breast carcinoma cells contain positive feedback loops to regulate multidrug resistance [82]. An array of functional experiments will be needed to confirm these data and further dissect the genes that regulate this feedback mechanism to elucidate if this is a global mechanism or specific.

Moving from QTL-to-Gene is hard

Traditional and multi-parent QTL mapping approached are useful in uncovering genomic regions harboring natural variation for phenotypes, but are notoriously incapable of elucidating true causative genes [32,71,83]. While GWAS in theory are capable of determining the causative site, many GWAS are severely underpowered to detect significant associations and struggle to

identify low frequency causative variants [15,17,84]. In this study, we successfully went from QTL-to-causative gene and likely to the causative variant within the gene employing QCTs, RNAi, over-expression, and CRISPR techniques. The group E [85] phase II enzyme, *Ugt86Dd* was shown to be functionally important for resistance to nicotine via RNAi and over-expression assays. Interestingly, *Ugt86Dd* is a candidate gene potentially involved in metabolism of a variety of compounds [32,86,87]. Thus, this gene may have a broad role in protecting *Drosophila* from xenobiotics. Upon checking the conservation of *Ugt86Dd*, a DIOPT analyze revealed *Ugt1A3* is the most likely human homolog, and is expressed in the brain, liver, and small intestine [88–90]. Within brain microsomes, *Ugt1A3* is induced by nicotine and likely glucuronidates nicotine, while glucuronidates nicotine in liver microsomes [89–91]. This conservation provides strong evidence that Ugt86Dd has a role with nicotine metabolism. During the course of this study a structural variant, a complex 22bp InDel, was discovered to reside in the second exon of *Ugt86Dd* and cause a premature stop codon. Over-expression studies demonstrated the 22bp InDel is a strong candidate variant. CRISPR studies demonstrated lab-derived variants, close to the natural variant, within *Ugt86Dd* were important for variation in nicotine resistance further supporting the notion that the complex 22bp InDel likely is a variant that provides variable effects to nicotine resistance in the DSPR.

An advantage of the DSPR and other multi-parental mapping populations is the ability to capture minor alleles within founders and increase the frequency of these minor alleles within the mapping population. One piece of data that demonstrates this nicely is the frequency of *Ugt86Dd* in the DSPR versus other mapping resources and natural populations. In the DSPR, *Ugt86Dd* deletion is maintained frequency of 0.401 in pA and 0.545 in pB (see founder haplotype data, Marriage *et al.*, (2014) while the DGRP harbors the *Ugt86Dd* deletion at a frequency of 0.035. To elucidate how common the *Ugt86Dd* InDel was in natural populations sampled across time,

latitude, and North America, we employed natural populations previously collected from orchards within Florida, Georgia, Maine, North Carolina, Pennsylvania, and South Carolina [47]. Raw sequence reads were retrieved from the NCBI SRA (BioProject accession PRJNA256231,[47]) for each population and analyzed for the presence of either the deletion or insertion allele of the InDel within *Ugt86Dd*. The results indicated that 9/14 populations harbored the deletion event with the lowest frequency at 1% and the highest frequency at 11%, the average population deletion frequency was ~0.02 (Supplementary Data 3.4). Thus, the *Ugt86Dd* deletion event is a rare recessive loss-of-function mutation captured in the DSPR.

Loss-of-function variants are predicted to be rare, deletious, and have causative roles in severe Mendelian diseases [92,93]. Recently, a whole-genome sequencing study demonstrated the human genome harbors ~ 100 loss-of-function variants with at least 30 in a homozygous state, and are likely a rich source of variation that contribute serve-to-mild effects on complex traits [94,95]. The 22bp deletion segregating in the DSPR within *Ugt86Dd* is a rare recessive loss-of-function mutation that likely provides variable effects on nicotine resistance. Unlike the predicted effects of loss-of-function variants, this loss-of-function variant within *Ugt86Dd* is not deletious to the organism neither under normal conditions nor completely under nicotine stress and contributes a modest effect to variable nicotine resistance. Thus, providing evidence in *D. melanogaster* that loss-of-function variants could be a rich source of genetic variation with modest effects on xenobiotic resistance similar to what is seen with transposable elements and CNV. In theory loss-of-function variants should be easy to discover, but loss-of-function variants are typically hard to elucidate due to alternative splicing-to-imperfect gene annotations. While alternative splicing is common in *D. melanogaster* to create a diverse set mRNAs from one gene leading to increased protein diversity [96], disrupted splicing patterns can create loss-of-function variants that provide variable effects on transcripts. However, these common loss-of-function

variants only affect a tiny fraction of the transcripts [94], and are masked by the majority of wildtype transcripts that are produced from a single gene [97]. Another problem are next-generation sequence data that generate short-reads. Short-read sequences that are unmappable for various reasons are filtered out, including reads that include loss-of-function variants [98]. Lastly, in the down-stream analysis, loss-of-function variants will be missed completely if the variant exist within pseudogenes that are excluded in genome-wide analysis [94]. With the soon to be available PacBio® sequence data for the DSPR, the short-read problem will eliminated leaving the possibility to uncover many loss-of-function variants involved in various complex phenotypes.

In summary, we utilized QCTs, RNAi, over-expression, and CRISPR to functional examine *Cyp28d1*, *Cyp28d2*, and *Ugt86Dd* to collectively determine which genes are functionally important and harbor natural variation for nicotine resistance between A3 and A4 in the DSPR. We determined that phase one enzymes, *Cyp28d1* and *Cyp28d2* likely harbor natural variation and are functionally important for resistance. Currently, no variants are known within or near these genes but with PacBio® sequence data available soon, variants will likely be uncovered. Over-expression and CRISPR experiments will need to be performed on these genes to determine if these gene are truly causative and to measure the effect each gene has on nicotine resistance within the DSPR. Our data firmly provides light on *Ugt86Dd*'s, a phase II enzyme, role in nicotine resistance. Our RNAi and over-expression data provides evidence that *Ugt86Dd* is functionally important in the midgut and malpighian tubules. In the course of this study, *Ugt86Dd* was found to harbor a natural 22bp InDel that segregates between A3 and A4 of the DSPR; moreover, segregates perfectly for nicotine resistance within the all DSPR founders. Over-expression experiments suggest this 22bp InDel is potentially the causative variant within *Ugt86Dd*. CRISPR experiments provide evidence that *Ugt86Dd* has a large effect in the DSRP

and the InDel is likely causative. Homologous recombination CRISPR experiments will be needed to determine the causative effects of the 22bp InDel and measure the genotypic contribution the InDel has on nicotine resistance within the DSPR.

Acknowledgements

We thank TRiP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks, the VDRC for providing transgenic RNAi fly stocks, and flygut (http://flygut.epfl.ch/) for donating a series of tissue specific GAL4 stocks. Moreover, we would like to thank FlyORF for donating pUASg.attB that was utilized in our over-expression studies. Lastly, we would like to thank BestGene for performing all injections utilized in this study and ACGT for Sanger sequencing our DNA products in this study.

Literature Cited

1. Glendinning JI. How do herbivorous insects cope with noxious secondary plant compounds in their diet? Entomol. Exp. Appl. 2002;104:15–25.

2. Kingsbury JM. Poisonous Plants of the United States and Canada. 3 edition. Prentice Hall; 1964.

3. Glendinning JI. How do predators cope with chemically defended foods? Biol. Bull. 2007;213:252–66.

4. Crow JF. Genetics of Insect Resistance to Chemicals. Annu. Rev. Entomol. 1957;2:227–46.

5. Ffrench-Constant RH. The Molecular Genetics of Insecticide Resistance. Genetics. 2013;194:807–15.

6. Xu C, Yong-Tao C, Kong A-NT. Induction of phase I, II and III drug metabolism/transport by xenobiotics. Arch Pharm Res. 2005;28:249–68.

7. Chung H, Sztal T, Pasricha S, Sridhar M, Batterham P, Daborn PJ. Characterization of Drosophila melanogaster cytochrome P450 genes. Proc. Natl. Acad. Sci. U. S. A. 2009;106:5731–6.

8. Guengerich FP. Cytochrome P450: what have we learned and what are the future issues? Drug Metab Rev. 2004;36:159–97.

9. Burchell B, Coughtrie MW. UDP-glucuronosyltransferases. Pharmacol. Ther. 1989;43:261–89.

10. You M, Yue Z, He W, Yang X, Yang G, Xie M, et al. A heterozygous moth genome provides insights into herbivory and detoxification. Nat. Genet. 2013;45:220–5.

11. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather AE, et al. Comprehensive Identification of Single Nucleotide Polymorphisms Associated with Beta-lactam Resistance within Pneumococcal Mosaic Genes. PLOS Genet. 2014;10:e1004547.

12. Wendler JP, Okombo J, Amato R, Miotto O, Kiara SM, Mwai L, et al. A Genome Wide Association Study of Plasmodium falciparum Susceptibility to 22 Antimalarial Drugs in Kenya. PLOS ONE. 2014;9:e96486.

13. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat. Genet. 2015;47:226–34.

14. Frick A, Suzuki OT, Benton C, Parks B, Fedoriw Y, Richards KL, et al. Identifying genes that mediate anthracyline toxicity in immune cells. Front. Pharmacol. [Internet]. 2015 [cited 2016 Jun 8];6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4398020/

15. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. 2001;69:124–37.

16. Spencer CCA, Su Z, Donnelly P, Marchini J. Designing Genome-Wide Association Studies: Sample Size, Power, Imputation, and the Choice of Genotyping Chip. PLOS Genet. 2009;5:e1000477.

17. Thornton KR, Foran AJ, Long AD. Properties and Modeling of GWAS when Complex Disease Risk Is Due to Non-Complementing, Deleterious Mutations in Genes of Large Effect. PLOS Genet. 2013;9:e1003258.

18. Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FH. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan Anopheles gambiae associated with resistance to DDT and pyrethroids. Insect Mol. Biol. 2000;9:491–7.

19. Defays R, Bertoli CI. Quantitative trait loci for response to ethanol in an intercontinental set of recombinant inbred lines of Drosophila melanogaster. Alcohol Fayettev. N. 2012;46:737–45.

20. Witzig TE, Laumann KM, Lacy MQ, Hayman SR, Dispenzieri A, Kumar S, et al. A phase III randomized trial of thalidomide plus zoledronic acid versus zoledronic acid alone in patients with asymptomatic multiple myeloma. Leukemia. 2013;27:220–5.

21. Reyes-Solis G del C, Saavedra-Rodriguez K, Suarez AF, Iv WCB. QTL Mapping of Genome Regions Controlling Temephos Resistance in Larvae of the Mosquito Aedes aegypti. PLOS Negl Trop Dis. 2014;8:e3177.

22. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, et al. A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 2009;5:e1000551.

23. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. Science. 2009;325:714–8.

24. Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS, et al. Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Res. 2011;21:1213–22.

25. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 2004;36:1133–7.

26. King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, et al. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 2012;22:1558–66.

27. Hall LM, Von Borstel RW, Osmond BC, Hoeltzli SD, Hudson TH. Genetic variants in an acetylcholine receptor from Drosophila melanogaster. FEBS Lett. 1978;95:243–8.

28. Carrillo R, Gibson G. Unusual genetic architecture of natural variation affecting drug resistance in Drosophila melanogaster. Genet Res. 2002;80:205–13.

29. Passador-Gurgel G, Hsieh W-P, Hunt P, Deighton N, Gibson G. Quantitative trait transcripts for nicotine resistance in Drosophila melanogaster. Nat Genet. 2007;39:264–8.

30. Li X, Bai S, Cass BN. Accord insertion in the 5' flanking region of CYP6G1 confers nicotine resistance in Drosophila melanogaster. Gene. 2012;502:1–8.

31. King EG, Macdonald SJ, Long AD. Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics. 2012;191:935–49.

32. Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. Genetics. 2014;198:45–57.

33. Bellen HJ, Levis RW, He Y, Carlson JW, Evans-Holm M, Bae E, et al. The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics. 2011;188:731–43.

34. Cook RK, Christensen SJ, Deal JA, Coburn RA, Deal ME, Gresens JM, et al. The generation of chromosomal deletions to provide extensive coverage and subdivision of the Drosophila melanogaster genome. Genome Biol. 2012;13:R21.

35. Ryder E, Ashburner M, Bautista-Llacer R, Drummond J, Webster J, Johnson G, et al. The DrosDel deletion collection: a Drosophila genomewide chromosomal deficiency resource. Genetics. 2007;177:615–29.

36. Parks AL, Cook KR, Belvin M, Dompe NA, Fawcett R, Huppert K, et al. Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. Nat. Genet. 2004;36:288–92.

37. Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature. 2007;448:151–6.

38. Buchon N, Osman D, David FPA, Fang HY, Boquete J-P, Deplancke B, et al. Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. Cell Rep. 2013;3:1725–38.

39. Long AD, Mullaney SL, Mackay T, Langley CH. Genetic Interactions between Naturally Occurring Alleles at Quantitative Trait Loci and Mutant Alleles at Candidate Loci Affecting Bristle Number in Drosophila Melanogaster. Genetics. 1996;144:1497–510.

40. Pasyukova EG, Vieira C, Mackay TFC. Deficiency Mapping of Quantitative Trait Loci Affecting Longevity in Drosophila melanogaster. Genetics. 2000;150:1129–46.

41. Bischof J, Björklund M, Furger E, Schertel C, Taipale J, Basler K. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Dev. Camb. Engl. 2013;140:2434–42.

42. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482:173–8.

43. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 2014;24:1193–208.

44. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics. 2014;196:961–71.

45. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 2013;8:2281–308.

46. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, et al. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics. 2013;194:1029–35.

47. Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. PLoS Genet. 2014;10:e1004775.

48. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 2008;9:477–85.

49. Mackay TFC. Quantitative Trait Loci In Drosophila. Nat. Rev. Genet. 2001;2:11–20.

50. Service PM. How good are quantitative complementation tests? Sci. Aging Knowl. Environ. SAGE KE. 2004;2004:pe13.

51. Suchail S, Debrauwer L, Belzunces LP. Metabolism of imidacloprid in Apis mellifera. Pest Manag. Sci. 2004;60:291–6.

52. Vasconcelos IM, Oliveira JTA. Antinutritional properties of plant lectins. Toxicon Off. J. Int. Soc. Toxinology. 2004;44:385–403.

53. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, et al. Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1. Genetics. 2007;175:1071–7.

54. Daborn PJ, Lumb C, Boey A, Wong W, Ffrench-Constant RH, Batterham P. Evaluating the insecticide resistance potential of eight Drosophila melanogaster cytochrome P450 genes by transgenic over-expression. Insect Biochem. Mol. Biol. 2007;37:512–9.

55. Daborn PJ, Lumb C, Harrop TWR, Blasetti A, Pasricha S, Morin S, et al. Using Drosophila melanogaster to validate metabolism-based insecticide resistance from insect pests. Insect Biochem. Mol. Biol. 2012;42:918–24.

56. Harrop TWR, Pearce SL, Daborn PJ, Batterham P. Whole-genome expression analysis in the third instar larval midgut of Drosophila melanogaster. G3 Bethesda Md. 2014;4:2197–205.

57. Hoi KK, Daborn PJ, Battlay P, Robin C, Batterham P, O'Hair RAJ, et al. Dissecting the Insect Metabolic Machinery Using Twin Ion Mass Spectrometry: A Single P450 Enzyme Metabolizing the Insecticide Imidacloprid in Vivo. Anal. Chem. 2014;86:3525–32.

58. Terra WR, Ferreira C. Insect digestive enzymes: properties, compartmentalization and function. Comp. Biochem. Physiol. Part B Comp. Biochem. 1994;109:1–62.

59. Ffrench-Constant RH, Mortlock DP, Shaffer CD, MacIntyre RJ, Roush RT. Molecular cloning and transformation of cyclodiene resistance in Drosophila: an invertebrate gamma-aminobutyric acid subtype A receptor locus. Proc. Natl. Acad. Sci. 1991;88:7209–13.

60. Ffrench-Constant RH. The molecular and population genetics of cyclodiene insecticide resistance. Insect Biochem. Mol. Biol. 1994;24:335–45.

61. Hu Y, Flockhart I, Vinayagam A, Bergwitz C, Berger B, Perrimon N, et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinformatics. 2011;12:357.

62. Vink JM, Smit AB, de Geus EJC, Sullivan P, Willemsen G, Hottenga J-J, et al. Genome-wide association study of smoking initiation and current smoking. Am. J. Hum. Genet. 2009;84:367–79.

63. Salleh MZ, Teh LK, Lee LS, Ismet RI, Patowary A, Joshi K, et al. Systematic Pharmacogenomics Analysis of a Malay Whole Genome: Proof of Concept for Personalized Medicine. PLOS ONE. 2013;8:e71554.

64. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, et al. A single p450 allele associated with insecticide resistance in Drosophila. Science. 2002;297:2253–6.

65. Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlötterer C. World-wide survey of an Accord insertion and its association with DDT resistance in Drosophila melanogaster. Mol. Ecol. 2004;13:2491–504.

66. Schlenke TA, Begun DJ. Strong selective sweep associated with a transposon insertion in Drosophila simulans. Proc. Natl. Acad. Sci. U. S. A. 2004;101:1626–31.

67. Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, O'Keefe M, et al. Cyp12a4 confers lufenuron resistance in a natural population of Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 2005;102:12807–12.

68. Marsano RM, Caizzi R, Moschetti R, Junakovic N. Evidence for a functional interaction between the Bari1 transposable element and the cytochrome P450 cyp12a4 gene in Drosophila melanogaster. Gene. 2005;357:122–8.

69. Chen S, Li X. Transposable elements are enriched within or in close proximity to xenobiotic-metabolizing cytochrome P450 genes. BMC Evol. Biol. 2007;7:46.

70. Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, et al. Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 Locus. Plos Genet. 2010;6:e1000998.

71. Najarro MA, Hackett JL, Smith BR, Highfill CA, King EG, Long AD, et al. Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. PLoS Genet. 2015;11:e1005663.

72. Wang J, Kean L, Yang J, Allan AK, Davies SA, Herzyk P, et al. Function-informed transcriptome analysis of Drosophila renal tubule. Genome Biol. 2004;5:R69.

73. Li H-M, Buczkowski G, Mittapalli O, Xie J, Wu J, Westerman R, et al. Transcriptomic profiles of Drosophila melanogaster third instar larval midgut and responses to oxidative stress. Insect Mol. Biol. 2008;17:325–39.

74. Neira Oviedo M, Vanekeris L, Corena-McLeod MDP, Linser PJ. A microarray-based analysis of transcriptional compartmentalization in the alimentary canal of Anopheles gambiae (Diptera: Culicidae) larvae. Insect Mol. Biol. 2008;17:61–72.

75. Pauchet Y, Wilkinson P, van Munster M, Augustin S, Pauron D, ffrench-Constant RH. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle Chrysomela tremulae reveals new gene families in Coleoptera. Insect Biochem. Mol. Biol. 2009;39:403–13.

76. Giraudo M, Unnithan GC, Goff GL, Feyereisen R. Regulation of cytochrome P450 expression in Drosophila: Genomic insights. Pestic Biochem Physiol. 2010;97:115–22.

77. Mittapalli O, Bai X, Mamidala P, Rajarapu SP, Bonello P, Herms DA. Tissue-Specific Transcriptomics of the Exotic Invasive Insect Pest Emerald Ash Borer (Agrilus planipennis). PLOS ONE. 2010;5:e13708.

78. Perry T, Batterham P, Daborn PJ. The biology of insecticidal activity and resistance. Insect Biochem Mol Biol. 2011;41:411–22.

79. Chahine S, O'Donnell MJ. Interactions between detoxification mechanisms and excretion in Malpighian tubules of Drosophila melanogaster. J. Exp. Biol. 2011;214:462–8.

80. Yang J, McCart C, Woods DJ, Terhzaz S, Greenwood KG, ffrench-Constant RH, et al. A Drosophila systems approach to xenobiotic metabolism. Physiol. Genomics. 2007;30:223–31.

81. Misra JR, Horner MA, Lam G, Thummel CS. Transcriptional regulation of xenobiotic detoxification in Drosophila. Genes Dev. 2011;25:1796–806.

82. Misra S, Ghatak S, Toole BP. Regulation of MDR1 expression and drug resistance by a positive feedback loop involving hyaluronan, phosphoinositide 3-kinase, and ErbB2. J. Biol. Chem. 2005;280:20310–5.

83. Rothammer S, Kremer PV, Bernau M, Fernandez-Figares I, Pfister-Schär J, Medugorac I, et al. Genome-wide QTL mapping of nine body composition and bone mineral density traits in pigs. Genet. Sel. Evol. 2014;46:68.

84. Long AD, Macdonald SJ, King EG. Dissecting complex traits using the Drosophila Synthetic Population Resource. Trends Genet. 2014;30:488–95.

85. Luque T, O'Reilly DR. Functional and phylogenetic analyses of a putative Drosophila melanogaster UDP-glycosyltransferase gene. Insect Biochem. Mol. Biol. 2002;32:1597–604.

86. Sun W, Margam VM, Sun L, Buczkowski G, Bennett GW, Schemerhorn B, et al. Genome-wide analysis of phenobarbital-inducible genes in Drosophila melanogaster. Insect Mol. Biol. 2006;15:455–64.

87. Mahapatra CT, Bond J, Rand DM, Rand MD. Identification of Methylmercury Tolerance Gene Candidates in Drosophila. Toxicol. Sci. 2010;116:225–38.

88. Strassburg CP, Strassburg A, Nguyen N, Li Q, Manns MP, Tukey RH. Regulation and function of family 1 and family 2 UDP-glucuronosyltransferase genes (UGT1A, UGT2B) in human oesophagus. Biochem. J. 1999;338 (Pt 2):489–98.

89. Kuehl GE, Murphy SE. N-glucuronidation of nicotine and cotinine by human liver microsomes and heterologously expressed UDP-glucuronosyltransferases. Drug Metab. Dispos. Biol. Fate Chem. 2003;31:1361–8.

90. Sakamoto M, Itoh T, Tukey RH, Fujiwara R. Nicotine regulates the expression of UDP-glucuronosyltransferase (UGT) in humanized UGT1 mouse brain. Drug Metab. Pharmacokinet. 2015;30:269–75.

91. Murphy CT, McCarroll SA, Bargmann CI, Fraser A, Kamath RS, Ahringer J, et al. Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. Nature. 2003;424:277–83.

92. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science. 2012;337:100–4.

93. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012;337:64–9.

94. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum. Mol. Genet. 2010;19:R125–30.

95. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012;335:823–8.

96. Venables JP, Tazi J, Juge F. Regulated functional alternative splicing in Drosophila. Nucleic Acids Res. 2012;40:1–10.

97. Kurmangaliyev YZ, Favorov AV, Osman NM, Lehmann K-V, Campo D, Salomon MP, et al. Natural variation of gene models in Drosophila melanogaster. BMC Genomics. 2015;16:198.

98. Metzker ML. Sequencing technologies - the next generation. Nat. Rev. Genet. 2010;11:31–46.

Figure 3.1

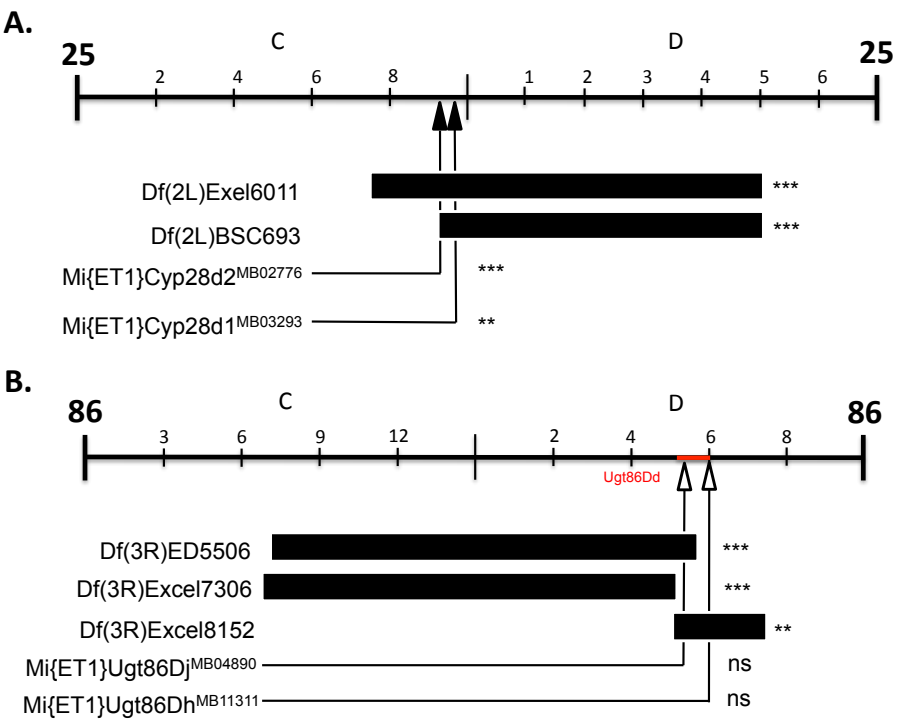

A.
25
C
D
25
2 4 6 8 1 2 3 4 5 6
Df(2L)Exel6011
***
Df(2L)BSC693
***
Mi{ET1}Cyp28d2MB02776
***
Mi{ET1}Cyp28d1MB03293
**
B.
86
C
D
86
3 6 9 12 2 4 6 8
Ugt86Dd
Df(3R)ED5506
***
Df(3R)Excel7306
***
Df(3R)Excel8152
**
Mi{ET1}Ugt86DjMB04890
ns
Mi{ET1}Ugt86DhMB11311
ns

Fig. 3.1. Fine mapping nicotine resistance QTL1 and QTL4 of the DSPR. Solid bars and triangles represent uncovered segregating allelic variation for nicotine resistance between A3 and A4. Open bars and triangles represent nonsignificant regions. Significance was assessed with asterisks represented by an ANOVA comparing resistance and susceptible genotypes for failure to complement (*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$). (A) Fine mapping of the 25C;25D nicotine resistance QTL1. Strains utilized were 5905, 6326, A3, A4, 23530, 23587, 7497, 26545. Long ticks mark sections and short ticks mark subsections of the 2L chromosome. Cytogenetic breakpoints studied induced in Df(2L)Excel6011 (25C8-25D5) and Df(2L)BSC693 (25C10-25D5). Additionally, insertional mutants Mi{ET1}Cyp28d1$^{[MB03293]}$ and Mi{ET1}Cyp28d2$^{[MB02776]}$ were studied. (B) Fine mapping of the 86C;86D nicotine resistance QTL4. Strains utilized were 5905, 6326, A3, A4, 9083, 24834, 27861, 7958, 7957, Long ticks mark sections and short ticks mark subsections of the 3R chromosome. Ugt86Dd is highlighted via red text, and the red bar depicts the 9 other UGT genes location under QTL4. of the Cytogenetic breakpoints studied induced in Df(3R)ED5506 (86C7-86D5) Df(3R)Exel7306 (86C7-86D5), and Df(3R)Exel8152 (86D5-86D7). Additionally, insertional mutants Mi{ET1}Ugt86Dj$^{[MB04890]}$ and Mi{ET1}Ugt86Dh$^{[MB11311]}$ were studied.

Figure 3.2

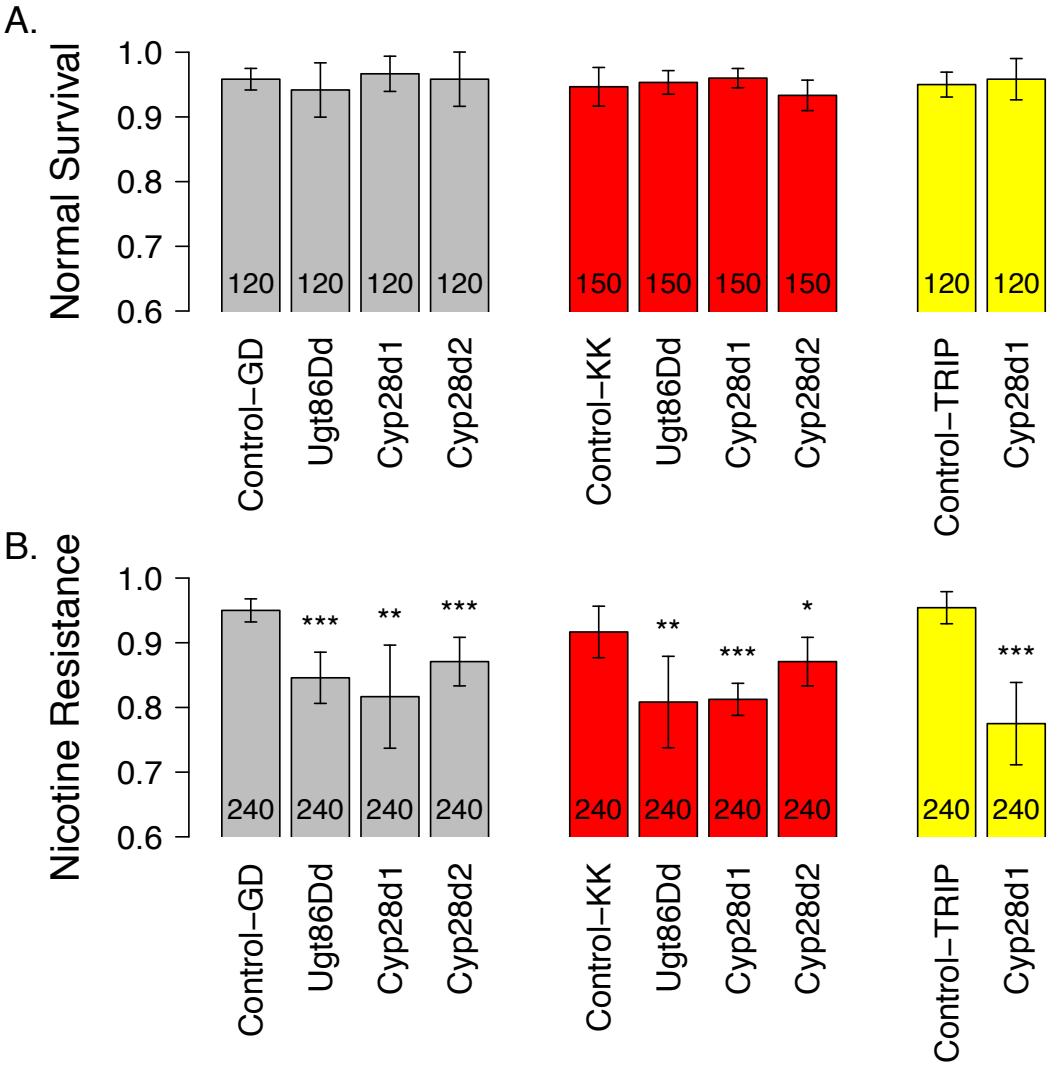

A.
Normal Survival
1.0
0.9
0.8
0.7
0.6
120
120
120
120
150
150
150
150
120
120
Control–GD
Ugt86Dd
Cyp28d1
Cyp28d2
Control–KK
Ugt86Dd
Cyp28d1
Cyp28d2
Control–TRIP
Cyp28d1
B.
Nicotine Resistance
1.0
0.9
0.8
0.7
0.6
***
**
***
**
***
*
***
240
240
240
240
240
240
240
240
240
240
Control–GD
Ugt86Dd
Cyp28d1
Cyp28d2
Control–KK
Ugt86Dd
Cyp28d1
Cyp28d2
Control–TRIP
Cyp28d1

Fig. 3.2. Effects of ubiquitous single gene RNAi depletion experiments. Ubiquitous driven GAL4-UAS RNAi of various candidate genes (*Cyp28d1*, *Cyp28d2*, and *Ugt86Dd*) tested on A) regular food or B) nicotine-laced food and compared to their respective control strains. Colored bars represent the specific RNAi library tested: gray = VDRC GD, red = VDRC KK, and yellow = TRiP. Strains utilized were 3954, 60000, 6016, 7868, 7870, 60100, 100353, 102626, 110259, 36304, 53892. Each bar depicts the mean nicotine resistance (± 1-SD) from a number of genetically identical F_1 1st instar larvae (sample size is at the bottom each bar). Significance is assessed with asterisks represented by a Welch's *t*-test comparing each RNAi genotype to its respective control (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Figure 3.3

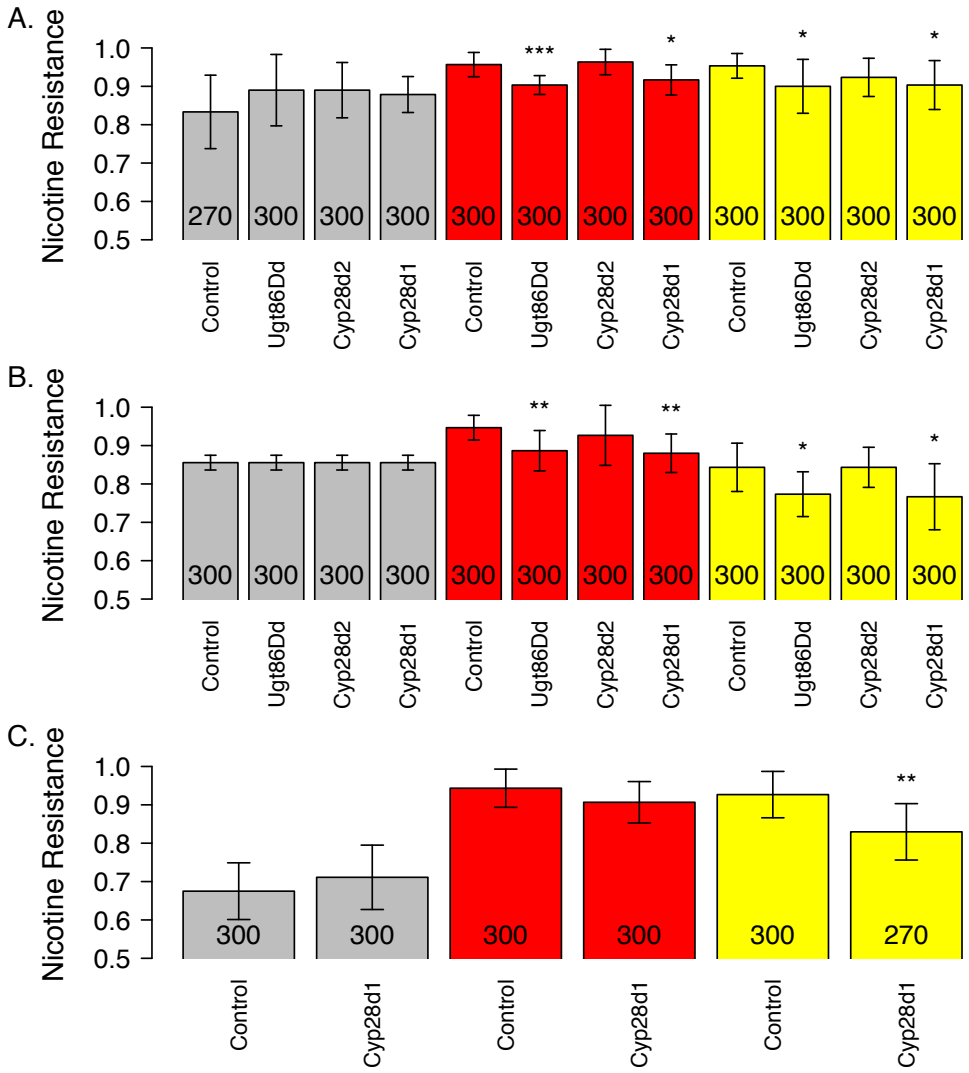

A.
Nicotine Resistance
1.0
0.9
0.8
0.7
0.6
0.5
270
300
300
300
300
300
300
300
300
300
300
300
***
*
*
*
Control
Ugt86Dd
Cyp28d2
Cyp28d1
Control
Ugt86Dd
Cyp28d2
Cyp28d1
Control
Ugt86Dd
Cyp28d2
Cyp28d1
B.
Nicotine Resistance
1.0
0.9
0.8
0.7
0.6
0.5
300
300
300
300
300
300
300
300
300
300
300
300
**
**
*
*
Control
Ugt86Dd
Cyp28d2
Cyp28d1
Control
Ugt86Dd
Cyp28d2
Cyp28d1
Control
Ugt86Dd
Cyp28d2
Cyp28d1
C.
Nicotine Resistance
1.0
0.9
0.8
0.7
0.6
0.5
300
300
300
300
300
270
**
Control
Cyp28d1
Control
Cyp28d1
Control
Cyp28d1

Fig. 3.3. Effects of tissue specific single gene RNAi depletion experiments. Tissue specific driven GAL4-UAS RNAi of various candidate genes (*Cyp28d1*, *Cyp28d2*, and *Ugt86Dd*) tested in our nicotine resistance assay against their respective control strains. Strains employed were 1967, 43656, l099, 60000, 6016, 7868, 7870, 60100, 100353, 102626, 110259, 36304, 53892. Colored bars represent the various tissue specific midgut Gal4 drivers: gray = posterior midgut (1967), red = anterior midgut (43656), and yellow = anterior midgut (l099). Each bar in the plot depicts the mean nicotine resistance (± 1-SD) from a number of genetically identical F_1 1st instar larvae (sample size is at the bottom each bar). Significance is assessed with asterisks represented by a Welch's *t*-test comparing each RNAi genotype to its respective control (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). (A) Depicts tissue specific RNAi performed using the VDRC-GD library. (B) Shows tissue specific RNAi performed using the VDRC-KK library. (C) Depicts tissue specific RNAi using the TRiP library.

Figure 3.4

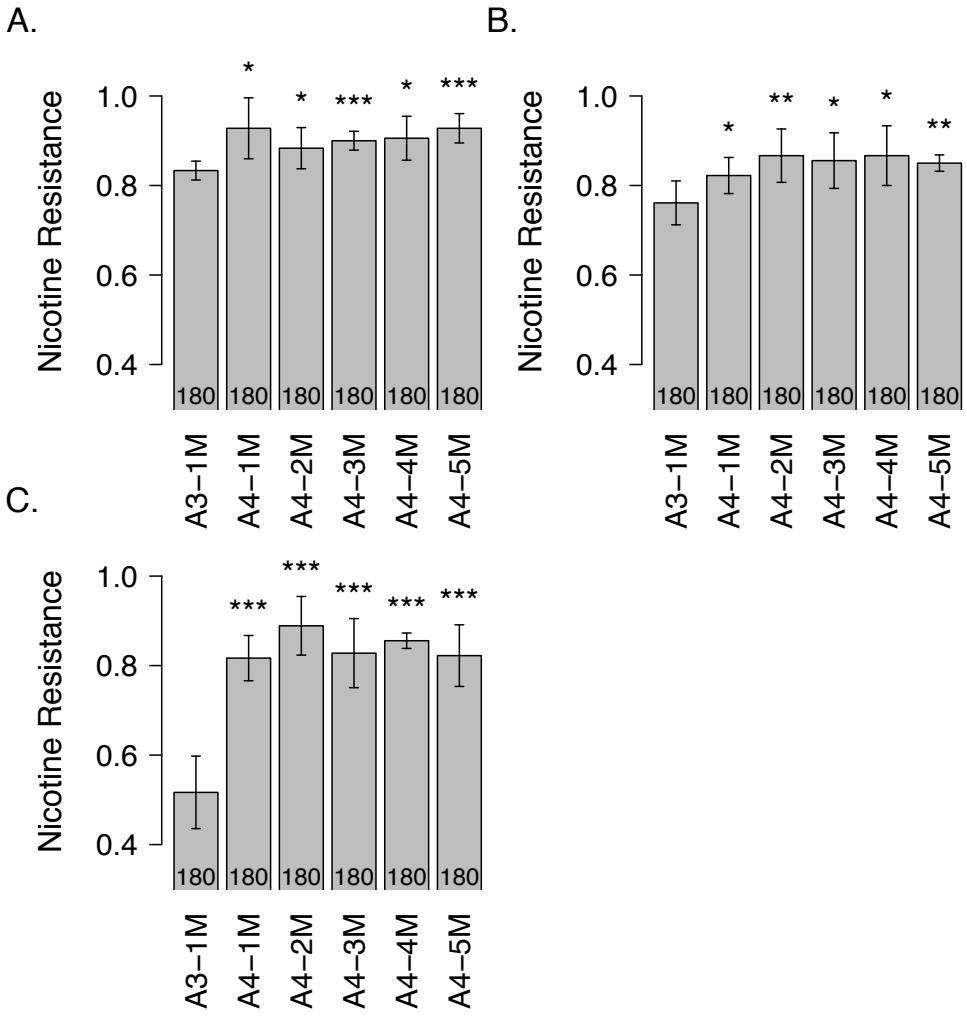

A.
B.
C.
Nicotine Resistance
1.0
0.8
0.6
0.4
180
A3–1M
A4–1M
A4–2M
A4–3M
A4–4M
A4–5M
*
**
***

Fig. 3.4. Effects of tissue specific single gene over-expression experiments. Tissue specific over-expression using the GAL4-UAS system to over-express the deletion or insert allele of *Ugt86Dd* to determine if the insert allele confers better nicotine resistance compared to the deletion allele. Strains employed were A3-1M, A4-1M, A4-2M, A4-3M, A4-4M, A4-5M , 1967, 43656, l099. Each bar in the plot depicts the mean nicotine resistance and ± 1-SD from a number of F_1 1st instar larvae (sample size is at the bottom each bar). Significance is assessed with asterisks represented by a Welch's *t*-test comparing each A4 genotype to the A3 genotype (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Figure 3.5

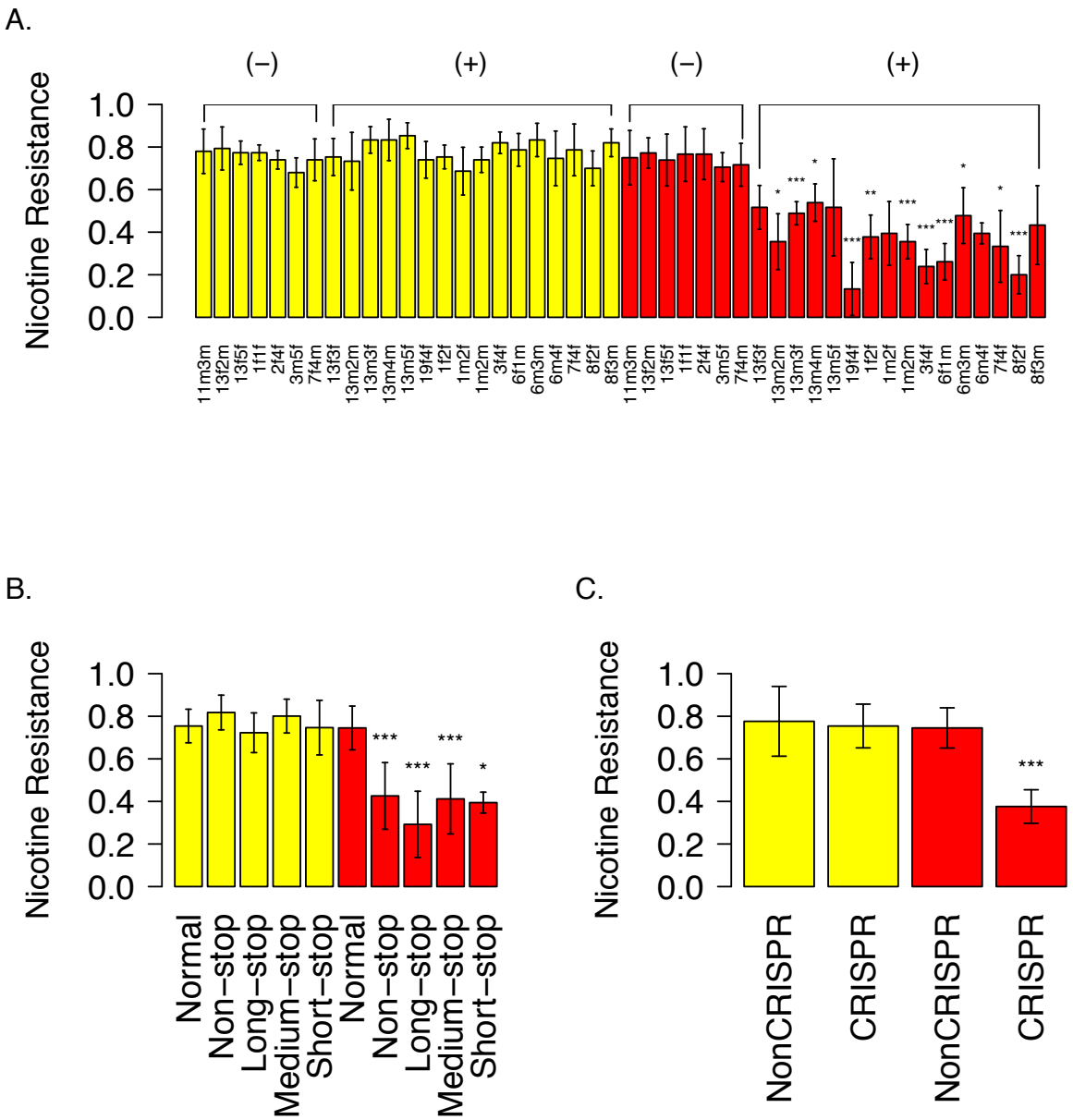

A.
(–)
(+)
(–)
(+)
Nicotine Resistance
1.0
0.8
0.6
0.4
0.2
0.0
11m3m
13f2m
13f5f
1f1f
2f4f
3m5f
7f4m
13f3f
13m2m
13m3f
13m4m
13m5f
19f4f
1f2f
1m2f
1m2m
3f4f
6f1m
6m3m
6m4f
7f4f
8f2f
8f3m
B.
Nicotine Resistance
Normal
Non–stop
Long–stop
Medium–stop
Short–stop
Normal
Non–stop
Long–stop
Medium–stop
Short–stop
C.
Nicotine Resistance
NonCRISPR
CRISPR
NonCRISPR
CRISPR

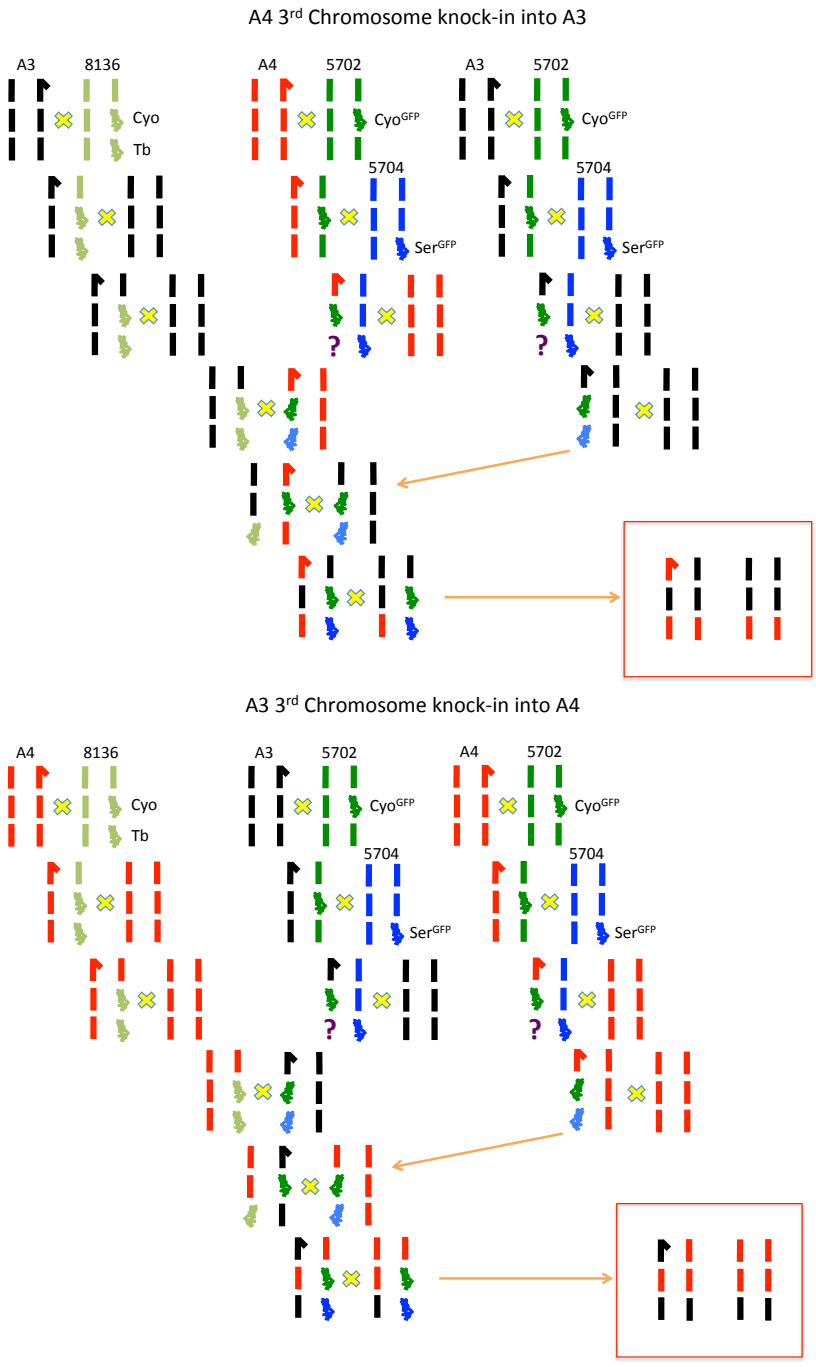
Fig. 3.5. Effects on *Ugt86Dd* CRISPR knockout. Each bar depicts the mean nicotine resistance (± 1-SD) for A) unique CRISPR or NonCRISPR genotypes and food, B) a number of CRISPR or NonCRISPR lines pooled together based on CRISPR status and stop codon, and C) by a number of CRISPR or NonCRISPR lines pooled together based on CRISPR status and food. A) Depicts unique CRISPR or NonCRISPR genotypes compared to food source. B) Depicts pooled CRISPR or NonCRISPR genotypes based on stop codon to food. C) Depicts comparison of pooled CRISPR or NonCRISPR genotypes to food source. An ANOVA or Welch's *t*-test was performed and significance was assessed with a Bonferroni correction and labeled with asterisks (*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$). A total of 5-6 replicates were tested on either control or nicotine-laced food for each CRISPR or NonCRISPR line and each replicate contained 30 1st instar larvae. Yellow represents control food, while red represents nicotine-laced food.

Table 3.1

Strain Number	Genotype	Source	Experiment
A3 (3844)	-	DSPR	QCT/Dom/RNAi/OE/CRISPR/Dominance
A4 (3852)	-	DSPR	QCT/Dom/RNAi/OE/CRISPR/Dominance
A3-Ugt86Dd_INS	A3/A4;A3,A4	Macdonald	Dominance
A4-Ugt86Dd_DEL/Ser(GFP)	A4/A3;A4;A3/TM3, P{GAL4-Hsp70.PB}TR2, P{UAS-GFP.Y}TR2, y+ Ser1	Macdonald	Dominance
8136	w[1118]; Sp-1/Cyo; Ly7/Tm6B	Bloomington	Dominance
5704	w^{1}; Sb^{1}/TM3, P{GAL4-Hsp70.PB}TR2, P{UAS-GFP.Y}TR2, y^{+} Ser^{1}	Bloomington	QCT/Dominance/RNAi
5702	w[1]; sna[Sco]/CyO, P{w[+mC]=GAL4-Hsp70.PB}TR1, P{w[+mC]=UAS-GFP.Y}TR1	Bloomington	QCT/Dominance/RNAi
5905	w[1118]	Bloomington	QCT
6326	w[1118]	Bloomington	QCT
9083	w[1118]; Df(3R)ED5506, P{w[+mW.Scer\FRT.hs3]=3'.RS5+3.3'}ED5506/TM6C, cu[1] Sb[1]	Bloomington	QCT
9083	w[1118]; Df(3R)ED5506, P{w[+mW.Scer\FRT.hs3]=3'.RS5+3.3'}ED5506/TM3, P{GAL4-Hsp70.PB}TR2, P{UAS-GFP.Y}TR2, y+ Ser1	DSPR	QCT
23530	w[1118]; Mi{ET1}Cyp28d1[MB03293]	Bloomington	QCT
23587	w[1118]; Mi{ET1}Cyp28d2[MB02776]	Bloomington	QCT
24834	w[1118]; Mi{ET1}Ugt86Dj[MB04890]	Bloomington	QCT
27861	w[1118]; Mi{ET1}Ugt86Dh[MB11311]	Bloomington	QCT
7497	w[1118]; Df(2L)Exel6011, P{w[+mC]=XP-U}Exel6011/CyO	Bloomington	QCT
7497	w[1118]; Df(2L)Exel6011, P{w[+mC]=XP-U}Exel6011/CyO, P{w[+mC]=GAL4-Hsp70.PB}TR1, P{w[+mC]=UAS-GFP.Y}TR1	DSPR	QCT
7957	w[1118]; Df(3R)Exel7306/TM6B, Tb[1]	Bloomington	QCT
7957	w[1118]; Df(3R)Exel7306/TM3, P{GAL4-Hsp70.PB}TR2, P{UAS-GFP.Y}TR2, y+ Ser1	DSPR	QCT
7958	w[1118]; Df(3R)Exel8152/TM6B, Tb[1]	Bloomington	QCT
7958	w[1118]; Df(3R)Exel8152/TM3, P{GAL4-Hsp70.PB}TR2, P{UAS-GFP.Y}TR2, y+ Ser1	DSPR	QCT
26545	w[1118]; Df(2L)BSC693, P+PBac{w[+mC]=XP3.WH3}BSC693/SM6a	Bloomington	QCT
26545	w[1118]; Df(2L)BSC693, P+PBac{w[+mC]=XP3.WH3}BSC693/CyO, P{w[+mC]=GAL4-Hsp70.PB}TR1, P{w[+mC]=UAS-GFP.Y}TR1	DSPR	QCT
60000	w[1118]	Bloomington	RNAi
7870	w[1118];;Cyp28d1-RNAi	Bloomington	RNAi
7868	w[1118];;Cyp28d2-RNAi	Bloomington	RNAi
6016	w[1118];;Ugt86Dd-RNAi	Bloomington	RNAi
60100	y,w[1118];P{attP,y[+],w[3`]	Bloomington	RNAi
100353	y,w[1118];P{attP,Ugt86Dd-RNAi}	Bloomington	RNAi
102626	y,w[1118];P{attP,Cyp28d2-RNAi}	Bloomington	RNAi
110259	y,w[1118];P{attP,;Cyp28d1-RNAi}/Cyo(YFP)	Bloomington	RNAi
110259	y,w[1118];P{attP,;Cyp28d1-RNAi}/Cyo	DSPR	RNAi
36304	y[1] v[1]; P{y[+t7.7]=CaryP}attP40	Bloomington	RNAi
53892	y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMJ21210}attP40	Bloomington	RNAi
3954	y[1] w[*]; P{w[+mC]=Act5C-GAL4}17bFO1/TM3, P{w[+mC]=GAL4-Hsp70.PB}TR2, P{w[+mC]=UAS-GFP.Y}TR2, y[+] Ser[1]	Bloomington	RNAi
3954	y[1] w[*]; P{w[+mC]=Act5C-GAL4}17bFO1/TM3, P{GAL4-Hsp70.PB}TR2, P{UAS-GFP.Y}TR2, y+ Ser1	DSPR	RNAi
43656	w[*]; P{w[+mC]=Scr-GAL4.4}1-3	Bloomington	Tissue Specfic RNAi, Tissue Specific OE
1967	y[1] w[*]; P{w[+mW.hs]=GawB}34B	Bloomington	Tissue Specfic RNAi, Tissue Specific OE
l099	-	Flygut	Tissue Specfic RNAi, Tissue Specific OE
C724	-	Dow	Tissue Specfic RNAi, Tissue Specific OE
C42	-	Dow	Tissue Specfic RNAi, Tissue Specific OE
C710	-	Dow	Tissue Specfic RNAi, Tissue Specific OE
uro	-	Dow	Tissue Specfic RNAi, Tissue Specific OE
6984	P{w[+mW.hs]=GawB}c754, w[1118]	Bloomington	Tissue Specfic RNAi
30828	w[*]; P{w[+mW.hs]=GawB}Aph-4[c232]	Bloomington	Tissue Specfic RNAi
30844	w[*]; P{w[+mW.hs]=GawB}c601[c601]	Bloomington	Tissue Specfic RNAi
24749	yw;;M{3xP3-RFP.attP}ZH-86Fb (with M{vas-int.Dm}ZH-2A)	BestGene	Tissue Specific OE
A3-1M	yw;;M{3xP3-RFP.attL{UAS:Ugt86Dd,w+}attR}ZH-86Fb	Macdonald	Tissue Specific OE
A4-1M	yw;;M{3xP3-RFP.attL{UAS:Ugt86Dd,w+}attR}ZH-86Fb	Macdonald	Tissue Specific OE
A4-2M	yw;;M{3xP3-RFP.attL{UAS:Ugt86Dd,w+}attR}ZH-86Fb	Macdonald	Tissue Specific OE
A4-3M	yw;;M{3xP3-RFP.attL{UAS:Ugt86Dd,w+}attR}ZH-86Fb	Macdonald	Tissue Specific OE
A4-4M	yw;;M{3xP3-RFP.attL{UAS:Ugt86Dd,w+}attR}ZH-86Fb	Macdonald	Tissue Specific OE
A4-5M	yw;;M{3xP3-RFP.attL{UAS:Ugt86Dd,w+}attR}ZH-86Fb	Macdonald	Tissue Specific OE
53892	y[1] M{vas-Cas9}ZH2A w[1118]	BestGene	CRISPR
53892	y[1] M{vas-Cas9}ZH2A w[1118]; ; A4	Macdonald	CRISPR

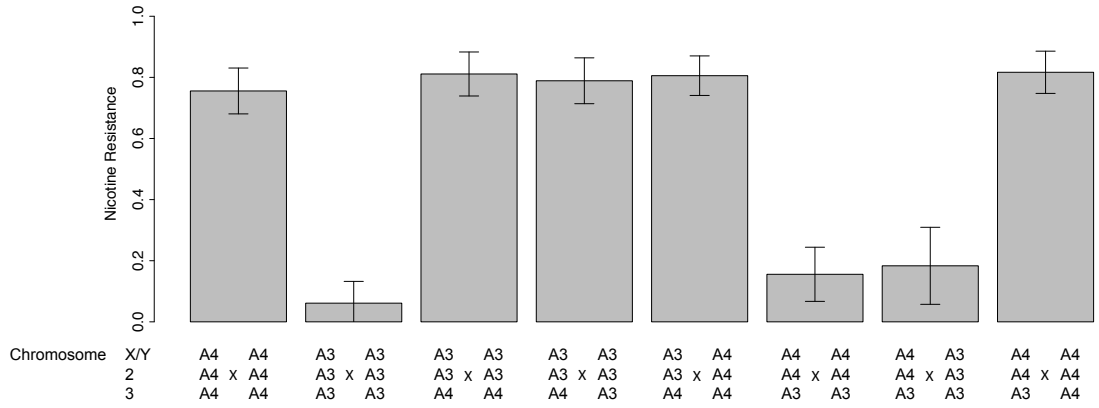
Supplementary Table 3.1. Strains utilized. The “Strain Number” column indicates the lines as called via BestGene, Bloomington Stock Center, DSPR, flygut, and VDRC. The “Genotype” column provides all known genotypes, while the Experiment column indicates the experiments these lines were utilized in.

Figure 3.S1


A4 3rd Chromosome knock-in into A3
A3
8136
Cyo
Tb
A4
5702
CyoGFP
5704
SerGFP
?
A3 3rd Chromosome knock-in into A4

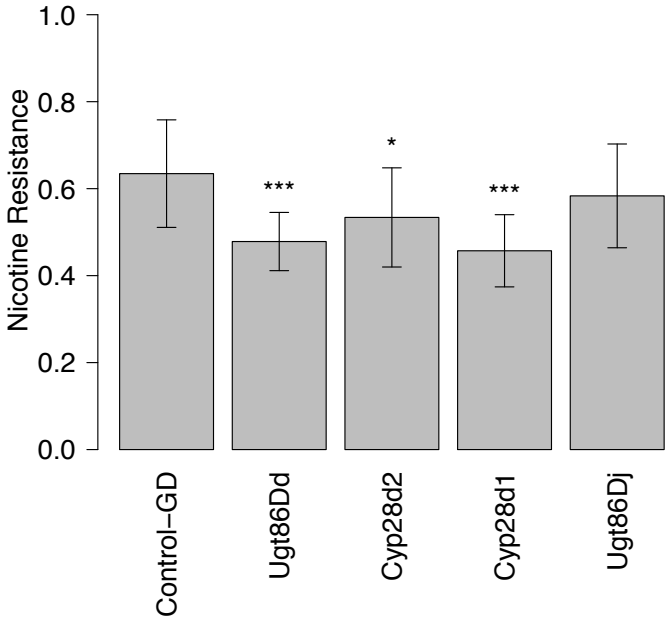
Supplementary Figure 3.1. Third chromosome knock-in between A3 and A4. Between A3 and A4 we swapped into each background the 3rd chromosome of the reciprocal strain. A schematic of the crossing scheme details each cross through the creation of the correct stocks that harbor the swapped 3rd chromosome.

Figure 3.S2


Nicotine Resistance
0.0
0.2
0.4
0.6
0.8
1.0
Chromosome
X/Y
2
3
A4 A4 x A4 A4
A4 A4 A4
A3 A3 x A3 A3
A3 A3 A3
A3 A3 x A3 A3
A4 A4 A4
A3 A3 x A3 A3
A4 A3 A3
A3 A4 x A4 A4
A4 A4 A4
A4 A4 x A4 A4
A3 A4 A4
A4 A3 x A3 A3
A3 A3 A3
A4 A4 x A4 A4
A3 A4 A4

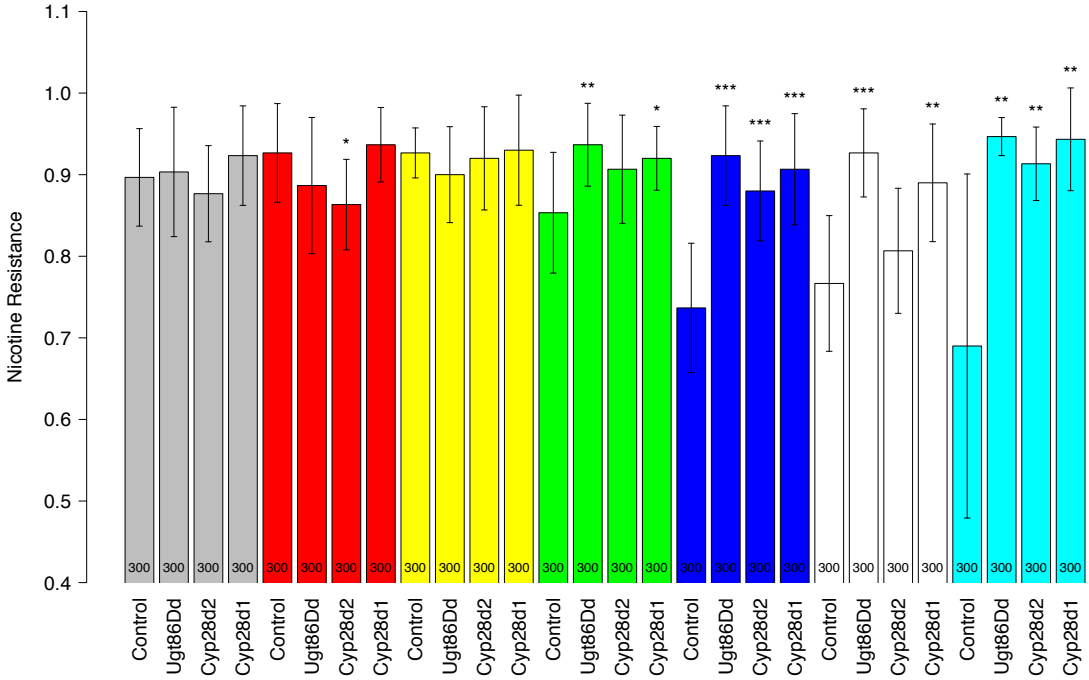
Supplementary Figure S2. Substituted third chromosome compared to A3 and A4. The third chromosome of A3 and A4 of the DSPR were swapped between each other to measure the effect the third chromosome has on nicotine resistance for each strain. A3-INS represents A3 with A4's version of *Ugt86Dd*, while A4-DEL represents A4 with A3's version of *Ugt86Dd* and is depicted below the genotype bars. Strains employed were A3, A4, A3-Ugt86Dd_INS, A4-Ugt86Dd_Del/Ser(GFP). Each bar depicts the mean nicotine resistance and a ± 1-SD from a number of genetically identical F_1 1st instar larvae. Thirty 1st instar larvae where in each test and 6 test per genotype where conducted.

Figure 3.S3


Nicotine Resistance
1.0
0.8
0.6
0.4
0.2
0.0
***
*
***
Control–GD
Ugt86Dd
Cyp28d2
Cyp28d1
Ugt86Dj

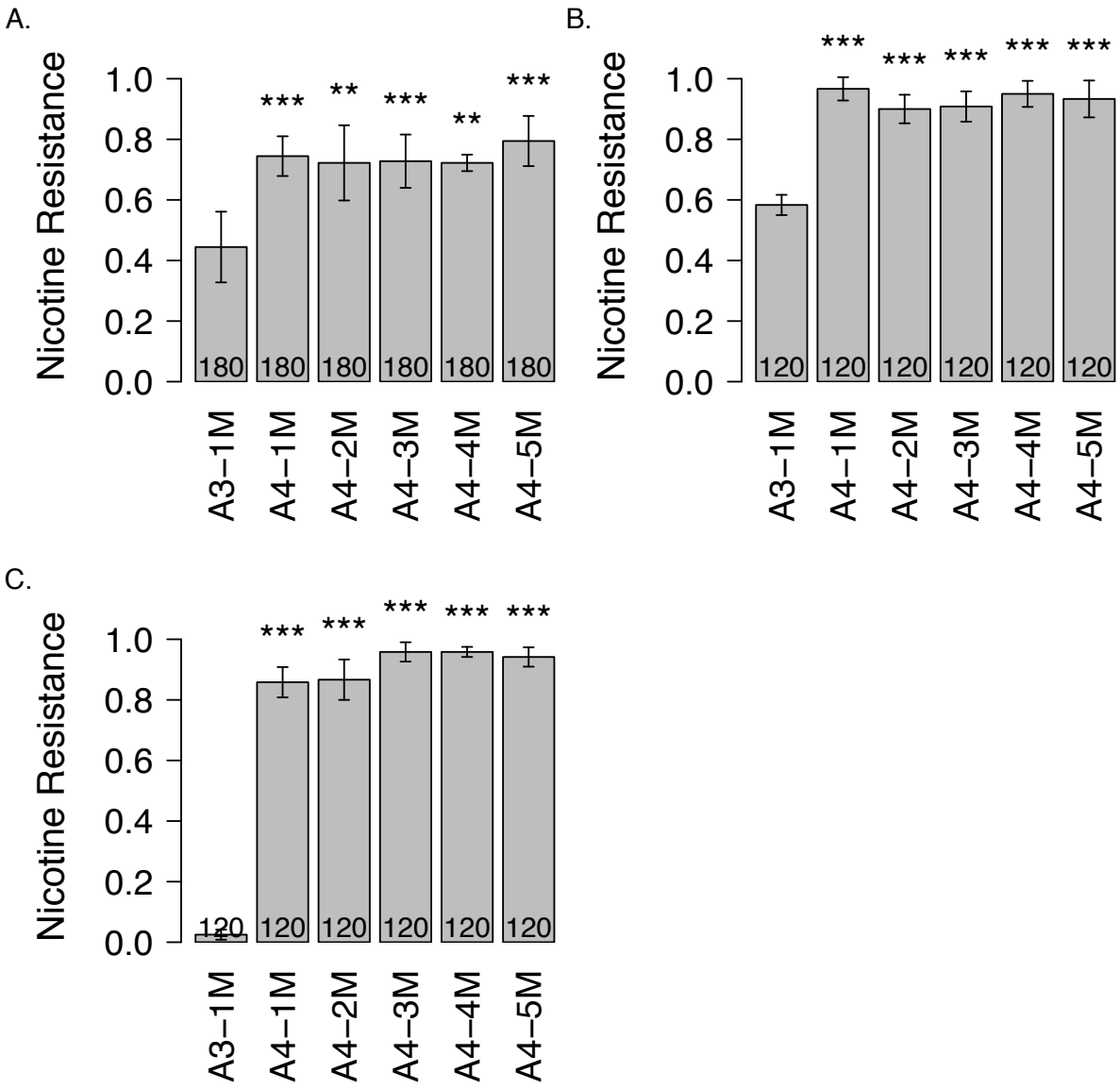
Supplementary Figure S3. Fraction of eclosed UAS-RNAi adults compared to control. We assessed the effects ubiquitous single gene RNAi knock-down has on nicotine resistance using the VDRC RNAi library (GD). A Ubiquitous driven GAL4 turned on UAS-RNAi for *Cyp28d1*, *Cyp28d2*, *Ugt86Dd*, and *Ugt86Dj*. Each bar depicts the mean nicotine resistance (± 1-SD) from the fraction of emerged adults that were genetically identical (GAL4-UAS RNAi) F_1 1st instar larvae. Significance is assessed with asterisks represented by a Welch's *t*-test comparing each RNAi genotype to its respective control (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Each Line was tested 15-16 times and each tested contained 30 1st instar larvae.

Figure 3.S4


Nicotine Resistance
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
300
Control
Ugt86Dd
Cyp28d2
Cyp28d1
*
**
***

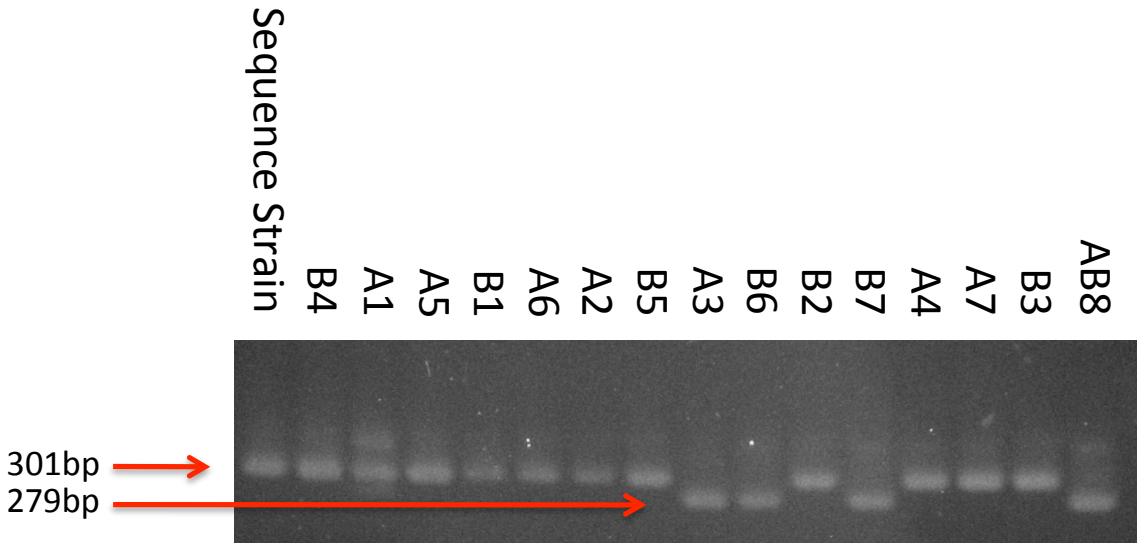
Supplementary S4. Fraction of eclosed RNAi adults compared to control. Tissue specific driven driven GAL4-UAS RNAi of various candidate genes (*Cyp28d1*, *Cyp28d2*, and *Ugt86Dd*) tested (6016, 7868, 7870, 100353, 102626, 110259, 53892)in our nicotine resistance assay against their respective control strains (60000, 6010, 36304). Colored bars represent the various tissue specific Gal4 drivers: gray = malpighian (30828), red = hindgut/malpighian (30844), yellow = fatbody (6984), green = malpighian tubule (C42), blue = malpighian tubule (C710), white = malpighian tubule (C724), and cyan = malpighian tubule (uro). Each bar depicts the mean nicotine resistance (± 1-SD) from a number of genetically identical F_1 1st instar larvae (sample size is at the bottom each bar). Significance is assessed with asterisks represented by a Welch's *t*-test comparing each RNAi genotype to its respective control (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Figure 3.S5


A.
Nicotine Resistance
1.0
0.8
0.6
0.4
0.2
0.0
***
**
***
**
***
180
180
180
180
180
180
A3−1M
A4−1M
A4−2M
A4−3M
A4−4M
A4−5M
B.
Nicotine Resistance
***
***
***
***
***
120
120
120
120
120
120
A3−1M
A4−1M
A4−2M
A4−3M
A4−4M
A4−5M
C.
Nicotine Resistance
***
***
***
***
***
120
120
120
120
120
120
A3−1M
A4−1M
A4−2M
A4−3M
A4−4M
A4−5M

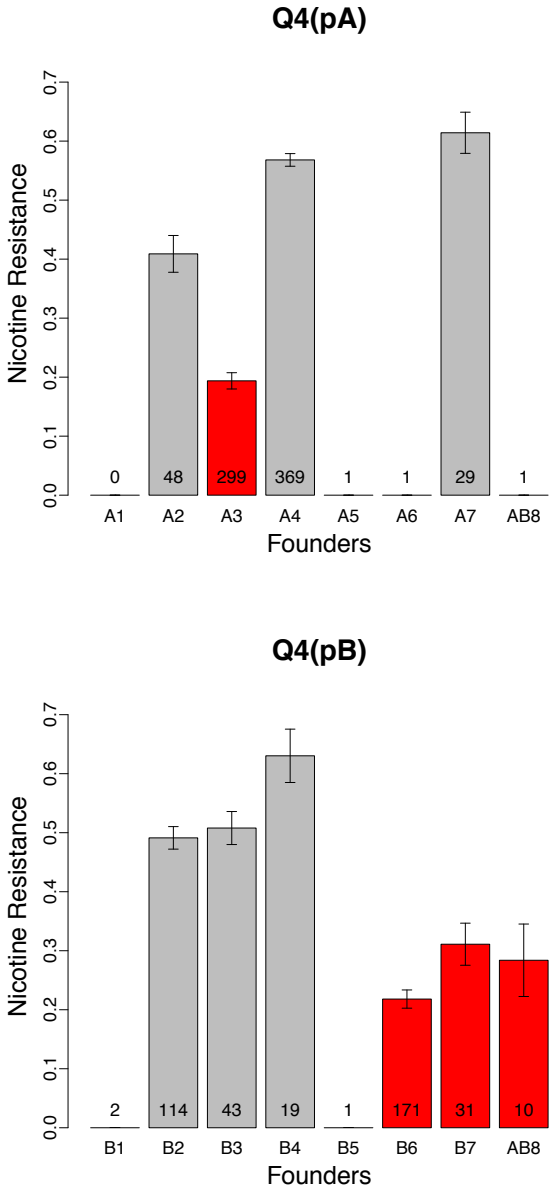
Supplementary Figure S5. Malpighian tubule *Ugt86Dd* over-expression experiments. Tissue specific over-expression using the GAL4-UAS system to over-express the deletion or insert allele of *Ugt86Dd* to determine if the insert allele confers better nicotine resistance compared to the deletion allele. Strains employed were A3-1M, A4-1M, A4-2M, A4-3M, A4-4M, A4-5M , c42, c710, c724, uro. Each bar in the plot depicts the mean nicotine resistance and ± 1-SD from a number of F_1 1st instar larvae (sample size is at the bottom each bar). Significance is assessed with asterisks represented by a Welch's *t*-test comparing each A4 genotype to the A3 genotype (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Figure 3.S6


Sequence Strain
B4
A1
A5
B1
A6
A2
B5
A3
B6
B2
B7
A4
A7
B3
AB8
301bp
279bp

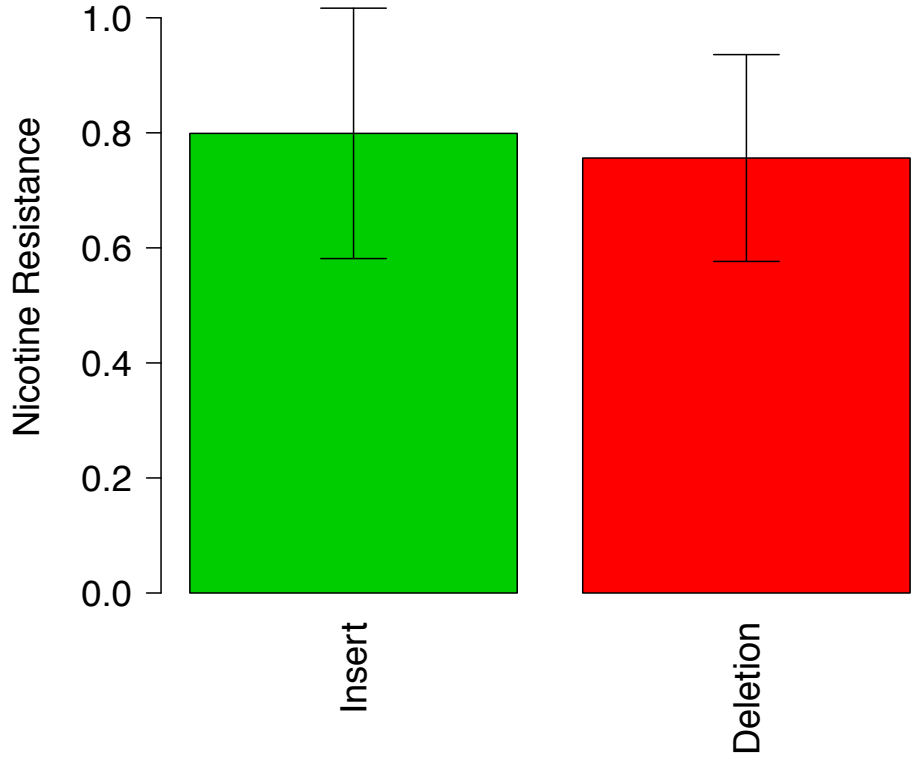
Supplementary Figure S6. Complex 22bp InDel is common in the DSPR. A 2% agarose gel demonstrating the 22bp segregating InDel variant within DSPR founders in this diagnostic PCR. A band of 301bp depicts the Insert loci of Ugt86Dd, while a band of 279bp depicts the deletion loci of Ugt86Dd.

Figure 3.S7


Q4(pA)
Nicotine Resistance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
48
299
369
1
1
29
1
A1
A2
A3
A4
A5
A6
A7
AB8
Founders
Q4(pB)
Nicotine Resistance
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
2
114
43
19
1
171
31
10
B1
B2
B3
B4
B5
B6
B7
AB8
Founders

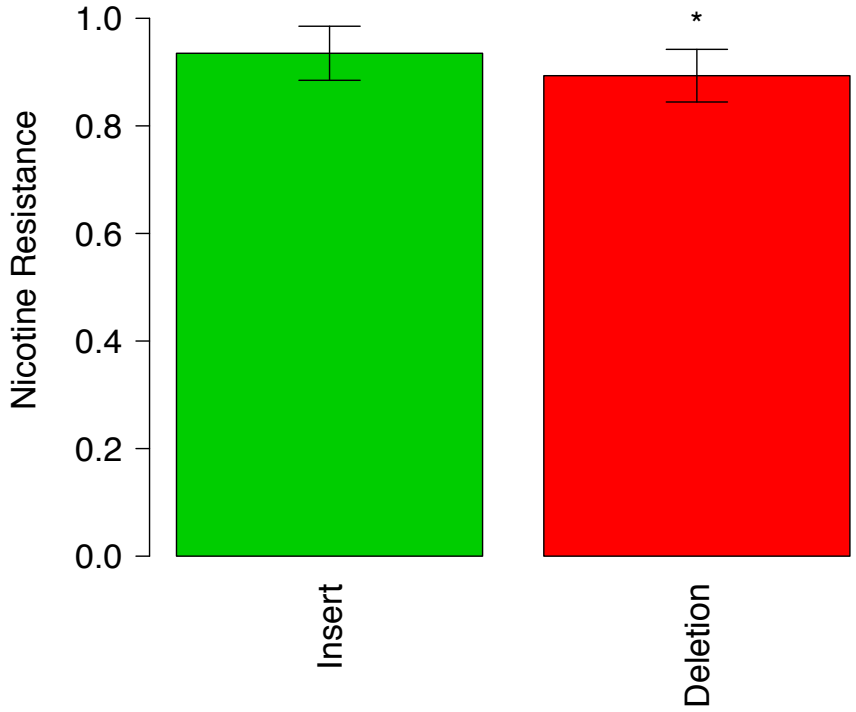
Supplementary Figure S7. Complex 22bp InDel indicates poor resistance. Founder haplotypes under QTL4 segregate for nicotine resistance based on 22bp InDel. Each bar are founder haplotype means and 1-SDs for QTL4 for each subpopulation of the DSPR. The number of founder genotypes assigned with a probability > 0.95 from RILs are present at the bottom of each bar, and founder means associated with at least 5 observations are only included. Red bars depict founders with the 22bp deletion, while gray bars are founders with the Insert allele.

Figure 3.S8



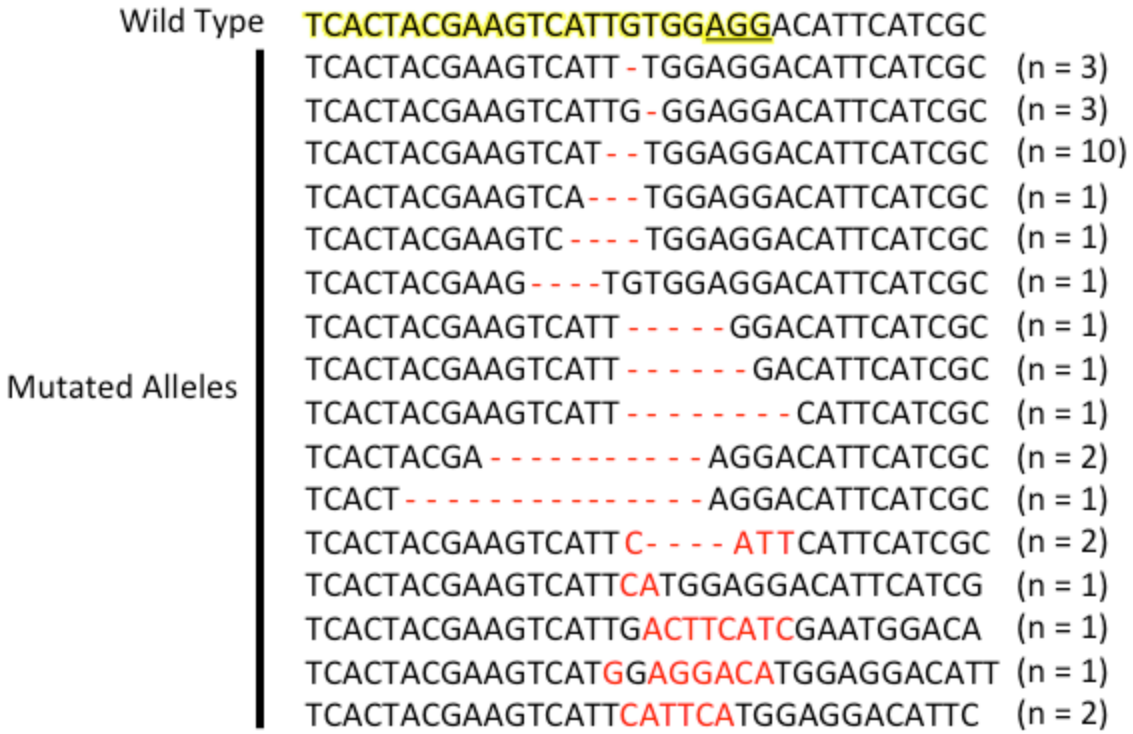
Supplementary Figure S8. DGRP lines associated with nicotine resistance based on InDel variant. Each bar depicts the mean nicotine resistance (± 1-SD) from a number of DGRP lines pooled together based on InDel variant. Significance was assessed with asterisks represented by a Welch's *t*-test comparing each pooled dataset (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). A total of 4 replicates were tested for each DGRP line selected and each replicate contained 30 1st instar larvae. Green represents DGRP lines pooled for the insert allele, and red represents DGRP lines pooled for the deletion allele.

Figure 3. S9



Supplementary Figure S9. Complex 22bp deletion likely functional. Each bar depicts the mean nicotine resistance (± 1-SD) from a number of 1st instar larvae that are not genetically unique but are sampled from the same population. Significance was assessed with asterisks represented by a Welch's *t*-test comparing each population (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). A total of 20 replicates were tested for population and each replicate contained 30 1st instar larvae. Green represents DGRP lines pooled for the insert allele, and red represents DGRP lines pooled for the deletion allele.

Figure 3.S10


Wild Type
TCACTACGAAGTCATTGTGGAGGACATTCATCGC
Mutated Alleles
TCACTACGAAGTCATT-TGGAGGACATTCATCGC (n = 3)
TCACTACGAAGTCATTG-GGAGGACATTCATCGC (n = 3)
TCACTACGAAGTCAT--TGGAGGACATTCATCGC (n = 10)
TCACTACGAAGTCA---TGGAGGACATTCATCGC (n = 1)
TCACTACGAAGTC----TGGAGGACATTCATCGC (n = 1)
TCACTACGAAG----TGTGGAGGACATTCATCGC (n = 1)
TCACTACGAAGTCATT-----GGACATTCATCGC (n = 1)
TCACTACGAAGTCATT------GACATTCATCGC (n = 1)
TCACTACGAAGTCATT--------CATTCATCGC (n = 1)
TCACTACGA----------AGGACATTCATCGC (n = 2)
TCACT----------AGGACATTCATCGC (n = 1)
TCACTACGAAGTCATTC----ATTCATTCATCGC (n = 2)
TCACTACGAAGTCATTCATGGAGGACATTCATCG (n = 1)
TCACTACGAAGTCATTGACTTCATCGAATGGACA (n = 1)
TCACTACGAAGTCATGGAGGACATGGAGGACATT (n = 1)
TCACTACGAAGTCATTCATTCATGGAGGACATTC (n = 2)

Supplementary Figure S10. Unique CRISPR mutants created. Wildtype *Ugt86Dd* sequence is shown up top. Yellow indicates the gRNA used and PAM sequence is underlined. All unique mutated alleles are shown below wildtype. Red dashes indicate various CRISPR induced mutations, while red letters indicate various CRISPR induced subsutitions and inserts. The number of times each unique CRISPR mediated mutation was discover is listed to the right of each unique mutated allele.

Supplementary Data 3.1. DGRP lines used to created fixed populations. The following 14 DGRP lines were differentially selected to start populations based on the Ugt86Dd InDel variant loci.

Deletion Population:

28199

25177

28185

28213

25206

25176

37525

Insert Population:

28239

25200

28160

28226

28197

25174

28295

Supplementary Data 3.2. CRISPR and NonCRISPR lines used for CRISPR experiments. The following 23 CRISPR or NonCRISPR lines were differentially selected to examine the effect a *Ugt86Dd* CRIPSR knockout has on nicotine resistance.

nonCRISPR lines:
3m5f
1m3m
1f1f
13f2m
7f4f
13f5f
2f4f

CRISPR lines:
6m4f
19f4f
19f2f
1m2f
8f3m
13m3f
6m3m
13m2m
13m4m
13m3f
3f4f
8f2f
6f1m
1m2m
13m5f
7f4f

Supplementary Data 3.3. *Ugt86Dd* CRISPR coding changes. For each unique CRISPR mutation, the amino acids sequence was analyzed and compared to the reference to determine what coding changes CRISPR introduced. Mutant amino acids are indicated in red and stop codons are represented via an x.

Ugt86Dd Coding/Polypeptide Sequence (Reference)

Red G is the 1bp missing in the sequenced strain Ugt86Dd allele (Leads to Glycine amino acid, although FlyBase has Alanine in protein sequence for download.)

Blue, yellow-highlighted section is the 22bp InDel.

ATGAGATTATTAACTGTGATCGCGATCGTGTTTTGCGCTTTGAGCGCCAAGCCGCTC
GAGTCGGAAAGTGCGAAGATT
M R L L T V I A I V F C A L S A K P L E S E S A K I
TTGGCTACGCTGCCGTTTCCGGGTCGATCTCAGTACATATTTGTGGAGAGTTACTTGA
AAGGCTTGGCAGCTAAGGGT
L A T L P F P G R S Q Y I F V E S Y L K G L A A K G
CATCAGGTGACCGTTATTAATGCCTTCAAGAACAAGGAGACACCAAATATGCGCTTT
ATTGAGGCCCTCAAAGCGCAC
H Q V T V I N A F K N K E T P N M R F I E A L K A H
GAATTCGCAGACGAGATGATGAGTTTGCTGAATGTGCCACTATTGTGGCAGCAGCTC
AATGCCATGGATTACATATTA
E F A D E M M S L L N V P L L W Q Q L N A M D Y I L
AACAAATTCATCGATGTAACTATGGAGGACGAGGGTGTCCAGAGGCTACTTAATTC
GGGAGAAACTTTCGATCTGGTG
N K F I D V T M E D E G V Q R L L N S G E T F D L V
CTGGCGGAGATGCTCCATATGGAACCGATGTACGCTTTTGCTCAGCATTTTAATGCC
ACTTTAGTGGGATTCTCCAGC
L A E M L H M E P M Y A F A Q H F N A T L V G F S S
TTTGGAACCGATAGAACTATTGATGAGGCTGCTGGAAATATATCACCCATATCATAT
AACCCGCTTGTGACCTCTCCT
F G T D R T I D E A A G N I S P I S Y N P L V T S P
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTGTG
GAGGACATTCATCGCCATTTC
R T D R M T F L E R L E N H Y E V I V E D I H R H F
GTTCACCTGCCACATATGAGAAATGTTTACAAAAAGTATTTCCCGAATGCAAAGAAA
ACCCTGGAGGAAGTCATGGAT

V H L P H M R N V Y K K Y F P N A K K T L E E V M D
AGTTTTTCGTTGATTCTGCTGGGCCAGCATTTTTCTTTGAGCTATCCGCGACCCTATT
TGCCCAACATGATTGAGGTT
S F S L I L L G Q H F S L S Y P R P Y L P N M I E V
GGTGGAATGCACATCTCGCACAAGCCGAAACCACTGCCAGAAGACATTAAACAATT
TATTGAGGGTTCGCCACATGGT
G G M H I S H K P K P L P E D I K Q F I E G S P H G
GTTATATACTTCTCCATGGGCTCCAATGTGAAGAGCAAGGATCTGCCACAGGAAACT
CGTGATACGCTGCTGAAGACC
V I Y F S M G S N V K S K D L P Q E T R D T L L K T
TTTGCCAAATTGAAGCAGAGAGTGCTGTGGAAATTCGAAGATGACGATATGCCTGG
AAAGCCAGCTAATGTGCTGATC
F A K L K Q R V L W K F E D D D M P G K P A N V L I
AAGAAATGGTATCCCCAGCCGGATATTCTGGCCCATCCGAATGTGAAGTTGTTCATC
AGCCACGGTGGTCTGCTGAGC
K K W Y P Q P D I L A H P N V K L F I S H G G L L S
AGCACCGAAAGCGTTTACTTTGGCAAACCCATACTGGGATTGCCGTGCTTCTATGAT
CAGCACATGAATGTGCAGCGC
S T E S V Y F G K P I L G L P C F Y D Q H M N V Q R
GCCCAGCGAGTGGGATTCGGCTTGGGCTTGGATCTGAATAATCTAAAGCAGGAGGA
TTTGGAAAAGGCCATTCAAACG
A Q R V G F G L G L D L N N L K Q E D L E K A I Q T
CTGCTCACTGATCCCAGTTATGCCAAAGCATCCTTGGCCATTTCCGAGCGGTATCGT
GATCAACCGCAATCAGCCGTC
L L T D P S Y A K A S L A I S E R Y R D Q P Q S A V
GATCGAGCTGTCTGGTGGACGGAGTACGTAATCAGGCACAATGGTGCTCCTCACCTG
CGAGCAACTTCCCGGGATCTC
D R A V W W T E Y V I R H N G A P H L R A T S R D L
AACTTCATTCAACTGAACAGCTTGGACACTTTAGCTGTGATACTGGCAGTACCTCTA
CTACTCGCTCTGTTAATCGTA
N F I Q L N S L D T L A V I L A V P L L L A L L I V
ACGTTATCTTGCAAATTATTGGGAGGAAAGAAACAGAAATGCTTACATGCTGATAA
GCTTAAGAAACATTAG
T L S C K L L G G K K Q K C L H A D K L K K H X

CRISPR mutant sequences

For each CRISPR lesion, list the mutation, the effect on the polypeptide, and show any mutant AA residues in red. Use "X" to represent stop codon.

3F.4F - 1bp (T) deleted, leads to STOP

CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTGGG
AGGACATTCATCGCCATTTCG
R T D R M T F L E R L E N H Y E V I G R T F I A I S
TTCACCTGCCACATATGA
F T C H I X

8F.3M - 1bp (G) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTTGG
AGGACATTCATCGCCATTTCG
R T D R M T F L E R L E N H Y E V I T R T F I A I S
TTCACCTGCCACATATGA
F T C H I X

1M.2F - 2bp (TG) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTGGA
GGACATTCATCGCCATTTCGT
R T D R M T F L E R L E N H Y E V I G G H S S P F R
TCACCTGCCACATATGAGAAATGTTTACAAAAAGTATTTCCCGAATGCAAAGAAAA
CCCTGGAGGAAGTCATGGATAG
S P A T T E K C L Q K V F P E C K E N P G G S H G X

6M.3M - 3bp (TTG) deleted, removes 1 AA, changes 1 AA
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATGGA
GGACATTCATCGCCATTTC
R T D R M T F L E R L E N H Y E V M E D I H R H F

7F.4F - 4bp (ATTG) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCTGGAG
GACATTCATCGCCATTTCGTTC
R T D R M T F L E R L E N H Y E V T R T F I A I S F
ACCTGCCACATATGA
T C H I X

13M.5F - 4bp (TCAT) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTGTGGAG
GACATTCATCGCCATTTCGTTC
R T D R M T F L E R L E N H Y E V W R T F I A I S F
ACCTGCCACATATGA
T C H I X

19F.4F - 5bp (GTGGA) deleted, leads to STOP

CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTGGA
CATTCATCGCCATTTCGTTCA
R T D R M T F L E R L E N H Y E V I G H S S P F R S
CCTGCCACATATGAGAAATGTTTACAAAAAGTATTTCCCGAATGCAAAGAAAACCCT
GGAGGAAGTCATGGATAG
P A T T E K C L Q K V F P E C K E N P G G S H G X

13M.4M - 6bp (GTGGAG) deleted, removes 2 AA
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTGAC
ATTCATCGCCATTTC
R T D R M T F L E R L E N H Y E V I D I H R H F

8F.2F - 8bp (GTGGAGGA) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTCAT
TCATCGCCATTTCGTTCACCT
R T D R M T F L E R L E N H Y E V I H S S P F R S P
GCCACATATGAGAAATGTTTACAAAAAGTATTTCCCGAATGCAAAGAAAACCCTGG
AGGAAGTCATGGATAG
A T T E K C L Q K V F P E C K E N P G G S H G X

13M.2M - 11bp (AGTCATTGTGG) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGGACATTC
ATCGCCATTTCGTTCACCTGCC
R T D R M T F L E R L E N H Y E G L S S P F R S P A
ACATATGAGAAATGTTTACAAAAAGTATTTCCCGAATGCAAAGAAAACCCTGGAGG
AAGTCATGGATAG
T T E K C L Q K V F P E C K E N P G G S H G X

6M.4F - 15bp (ACGAAGTCATTGTGG) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTAG
R T D R M T F L E R L E N H X

1F.2F - Overall 1bp insert (G->CA), leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTCAT
GGAGGACATTCATCGCCATTT
R T D R M T F L E R L E N H Y E V I H G G H S S P F
CGTTCACCTGCCACATATGAGAAATGTTTACAAAAAGTATTTCCCGAATGCAAAGAA
AACCCTGGAGGAAGTCATGGA
R S P A T T E K C L Q K V F P E C K E N P G G S H G
TAG

X

13F.3F - Overall 5bp insert (G->CATTCA), leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTCAT
TCATGGAGGACATTCATCGCC
R T D R M T F L E R L E N H Y E V I H S W R T F I A
ATTTCGTTCACCTGCCACATATGA
I S F T C H I X

6F.1M - Overall 6bp insert (TG->GGAGGACA), adds 2 AA, changes 2 AA
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATGGA
GGACATGGAGGACATTCATCGC
R T D R M T F L E R L E N H Y E V M E D M E D I H R

1M.2M - Overall 8bp insert (TGGA->ACTTCATCGAAT), leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTGAC
TTCATCGAATGGACATTCATC
R T D R M T F L E R L E N H Y E V I D F I E W T F I
GCCATTTCGTTCACCTGCCACATATGA
A I S F T C H I X

13M.3F - Overall 4bp deleted (GTGGAGGA->CATT), leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTCAT
TCATTCATCGCCATTTCGTTC
R T D R M T F L E R L E N H Y E V I H S F I A I S F
ACCTGCCACATATGA
T C H I X

DSPR mutant sequence

A3.UGT86Dd - 22bp (CATTGTGGAGGACATTCATCGC) deleted, leads to STOP
CGAACCGATCGGATGACCTTTCTGGAGCGCTTGGAAAATCACTACGAAGTCATTTCG
TTCACCTGCCACATATGA

R T D R M T F L E R L E N H Y E V N S F T C H I X

Supplement Data 3.4. Frequency of *Ugt86Dd* InDel variant in natural populations.

Population	Insert Allele	Deletion Allele	Frequency
Florida (rep 1)	55	2	0.04
Florida (rep 2)	32	0	0.00
Georgia	99	1	0.01
South Carolina	98	2	0.02
North Carolina	20	1	0.05
Maine (rep 1)	60	0	0.00
Maine (rep 2)	15	0	0.00
Pennsylvania	172	0	0.00
Pennsylvania	61	0	0.00
Pennsylvania	24	3	0.11
Pennsylvania	78	1	0.01
Pennsylvania	63	1	0.02
Pennsylvania	72	5	0.06
Pennsylvania	51	3	0.06

Chapter IV

Conclusions and Future Directions

Section 4.1 Concluding Remarks

Many biological scientist study traits that are thought to monogenetic disorders with large effects such as cystic fibrous; however, it is increasing clear that these traits are not so simple in nature but are actually complex in nature [1,2]. Due to the complexity of life, it is not surprising a large number of traits are polygenic in nature. Moreover, underpinning the genetic factors contributing to complex traits has been challenging due to the genetic architecture of the genome, environmental factors, and various interactions acting on the trait. Many traits have been studied extensively and researchers can only draw broad correlations about the trait. However, little attention has been paid in discovering the variable genetic factors affecting an array of complex traits in natural populations. Thus, we sought to better understand the genetic factors that contribute to longevity and xenobiotic resistance using *D. melanogaster* as a study model. This dissertation provided insights on variation on mated female lifespan and nicotine drug resistance within *D. melanogaster*. Both traits are multifactorial in nature and was/is likely affected via direct and indirect natural selection. The ability to live longer is likely to increase the number of progeny one can propagate, while withstanding the toxic effects of various xenobiotics likely opens up niches *D. melanogaster* can live and thrive, and also the ability to survive and propagate more progeny in the nature.

Variation in lifespan has been observed time and time again within species and between species [3–6]. Thus, it is of great interest to tease apart the genetic factors that create this variation. Previously, researchers have employed Quantitative Trait Loci (QTL) mapping to localize genomic regions harboring this genetic variation for lifespan within *D. melanogaster* [7–19]. With success, several studies have mapped lifespan QTL, but to broad regions of the genome containing thousands of genes leaving the causative gene(s) elusive. To more thoroughly

investigate the genetic variation controlling lifespan, QTL were mapped for lifespan in the DSPR. The architectural design of the DSPR allows simultaneous estimates of QTL effect size, QTL frequency, offers a plethora of segregating variation, and statistical power to detect QTL with small effects [20,21].

Chapter II revealed that 98.2% of the genome has been mapped for lifespan when inspecting our mapped QTL and comparing to the fields mapping data. This could indicate that perhaps most genes influence lifespan to some degree. Moreover, this also implies the genetic architecture of lifespan is more complex in *D. melanogaster* than previously thought and any genetic factors contributing to variation in lifespan within populations are probably small and subtle. Thus, any QTL mapped with subtle effect genes probably are composed of several tightly linked genes (gene clusters) that drive the phenotype in the same direction, enabling us to detect them via QTL mapping. Moreover, our QTL are a minimal set uncovered in the DSPR and there are numerous explanations for this. One reason to miss QTL is due to a cluster of genes around a marker that affect lifespan in opposite directions leaving the appearance the marker found no QTL. In order to address this problem, a higher sample size of RILs than the 805 RILs employed in this study will be needed to find even smaller effect QTL and/or increase the amount of segregating variation. By creating a population with 8 inbred founders, as the DSPR did, we greatly limit the identification of QTL in only a small subset of the genetic backgrounds nature offers, in comparison to the infinite number of genetic backgrounds obtainable in nature. Lastly, some of these opposing effects could be linked to sex, and doing sex specific mapping in the DSPR could uncover these genes with subtle sex specific effects.

Data collected via RNAseq and GO analysis revealed similar results reported previously [17,22–30]. Specifically, defense response genes are consistently up-regulated in older age and may have a role in partially determining variation in lifespan in natural populations. The

mechanism that manifest variable effects in lifespan through defense response genes is very poorly understood. Two ideas have been postulated about why defense response genes are up-regulated in older age: 1) direct effect of continuous pathogen exposure has no effect on lifespan or 2) loss of transcriptional regulation of these genes leads to costly effects that cause premature death [30–32].

Supporting the second idea mentioned above, our genome-wide RNAseq data uncovered several microbial recognition receptors, such as the peptidoglycan recognition proteins (PGRPs) and antimicrobials, (i.e., attacins and diptericins). These receptors are differentially expressed between young and old, short and long lived mated females, demonstrating the immune deficiency (Imd) pathway could have variable effects for lifespan within the DSPR. Interestingly, an inflammation state created by long-term activation via the Imd pathway is physiologically costly and reduces lifespan [33–35]. Our data supports a model where loss of transcriptional regulation of the Imd pathway induces a physiologically costly inflammation state that leads to variation in lifespan. Furthermore, the key NF-κB-like transcription factor, *Relish*, that activates the array of antimicrobial peptides of the Imd pathway is differentially expressed and is under our mapped QTL. Inappropriate activation of *Relish* can have adverse biological effects and reduce lifespan [34,36]. Thus, this is an excellent candidate gene to examine for variants that have variable effects on lifespan.

The Janus kinase and signal transducer and activator of transcription (JAK/STAT) pathway has recently been implicated in variation in lifespan via systemic expression of the JAK/STAT pathway leading to shorter lifespan, enhanced fat storage, increased insulin resistance, and hyperglycemia [37]. Our genome-wide RNAseq data uncovered the main transmembrane signal-transducing receptor, *Domeless*, required for all JAK/STAT interactions [38] was differentially expressed, with significantly higher expression in short-lived old females

in comparison to all other contrast. Moreover, we also observed a much higher expression of differentially expressed Turandot genes (*TotA*, *TotB*, *TotC*, and *TotM*) in short-lived old females, indicating a possible environmental and/or bacterial stress. Another plausible explanation is the loss of transcriptional regulation of the JAK/STAT pathway is occurring in the short-lived old females inducing physiologically costly stress responses, similar to what is seen in the Imd pathway. Just as inappropriate expression of *Relish* can have adverse biological effects and reduce lifespan. Also, inappropriate expression of *Domeless* could have these similar effects.

In nature, agriculture businesses generally apply insecticides/pesticides to fend off pest such as insects. For a while these pest are susceptible to the insecticides/pesticides, but over time these insects continually become resistance to those insecticides/pesticides. A great example is the Aedes mosquitoes that are located in various areas of the Americas and are the carrier of the Zika virus [39]. These mosquitoes are highly resistant to Permethrins chemicals that are employed in the Americas to kill crop pest [39]. Thus, because of this resistance, scientist are having great difficulty in battling the Zika virus. It is clear that resistance development is not only an agriculture problem but a public health problem. This provides a strong assertion that more resources need to be dedicated in dissecting xenobiotic resistance. Currently, the scientific community has a broad understanding how xenobiotics are broken down [40], but it is unknown what genes are involved in specific xenobiotic detoxification, For example, how variation in these genes provides variable effects to the detoxification process. Thus, it is critical to understand the genetic basis [41] of xenobiotic resistance in insects to give general insights to the genetic factors that influence the variable metabolism of drugs. For example, phenotypic variation for nicotine resistance has been demonstrated in a variety of genetically different *Drosophila* strains [42–45]. We followed up on Marriage *et al.* (2014) QTL mapping data by functionally validating *Cyp28d1* and *Cyp28d2*, and *Ugt86Dd* via QCTs, RNAi, over-expression,

and CRISPR. Finding from these approaches uncovered genes that are utilized in the metabolism of nicotine and will help understand the variable effects of nicotine metabolism. Moreover, these results will help the scientific community understand the variable genetic effects of xenobiotic resistance.

Currently GWAS and QTL mapping are widely employed in an attempt to localize genes and/or variants associated with xenobiotic resistance. However, each of these approaches often have trouble localizing genes and/or variants associated with the resistance. GWAS offer great resolution, but often utilizes small samples size and lacks power [46], and only uncovers common variants leaving rare variants essentially invisible [47]. QTL mapping offers strong power but has poor resolution [48]. To maintain the strengths of each of the approaches to better attempt the ability to localize genes and/or variants a series of multi-parental mapping panels have been developed [20,49–51]. However, even with the improved resolution, localization of true causative gene(s) and/or variant(s) associated with the phenotype remains challenging. This improved approach will instead highlight a small number of candidate genes, at least in *Drosophila*, that are LD with the true gene(s) and/or variant(s) [45,52,53]. Chapter III provides an inherently elegant way to get from QTL to gene by employing array of functional assays that include QCTs, RNAi, over-expression, and CRISPR to functionally validate candidate genes uncovered in a nicotine resistance screen by Marriage *et al.*, (2014).

Chapter III revealed *Cyp28d1* and *Cyp28d2* appear to be functionally important for nicotine resistance, while *Ugt86Dd* is a causative nicotine resistance gene. Moreover, we have mounting evidence demonstrating a complex 22bp deletion within *Ugt86Dd* is contributing variable nicotine resistance effects in the DSPR. A DIOPT homology analyze uncovered that *Ugt1A3* is the likely human/mice homology of *Ugt86Dd*. Recently, *Ugt1A3* was shown to be upregulated in presence of nicotine and perform glucuronidation reactions on nicotine in the

brain and small intestine [54,55] providing a conserved functional role from *Drosophila*-to-human. However, it is not well understood what genes are involved in nicotine metabolism across taxa. Moreover, even within *D. melanogaster* very disparate results have been uncovered with the only common finding between all studies is that phenotypic variation for nicotine resistance exist in natural populations. These disparate finding demonstrate the notion that nicotine resistance is highly polygenic. Interestingly, our data suggest that *Ugt86Dd* has subtle to modest effects on the phenotype based on background. This supports the idea other variants are involved in resistance and can mask the effects of *Ugt86Dd* has on nicotine resistance. A quick KEGG search predicts *Ugt86Dd* is likely involved in nicotine metabolism via one of the many predicted pathways, further supporting the role of other gene variants. Currently, the only other gene that has been shown to have functional role in nicotine resistance is the phase I enzyme *Cyp6g1* [56]. Thus, we have a phase I and phase II enzymes implicated in nicotine resistance and likely two more phase I enzymes *Cyp28d1* and *Cyp28d2*. Further examination needs to determine the causative effect of *Cyp28d1* and *Cyp28d2*.

Loss of function variants are typically associated with large effect disease-causing phenotypes, such as cystic fibrosis. This disease is the most common deadly genetic disease within the Caucasian population and is caused from a homozygous copy of the loss of function cystic fibrosis transmembrane conductance regulator gene [57]. It is established that loss of function of this regulator gene has severe-to-neutral effects on the phenotype [57–59]. Recently, large-scale resequencing have revealed the human genome harbors ~100 loss of function variants per genome of which ~30 are in a homozygous state [60]. These variants are typically found to be at low frequencies in genome via purifying selection and likely are mildly deleterious [60]. Thus, loss of function variants are an excellent source of genetic factors that create variable effects for complex traits. For instance, pharmacogenetic studies have demonstrated several

times that loss of function variants create variable effects on drug metabolism [61,62]. With the amount of conservation between the *D. melanogaster* genome and human genome, it would not be a surprise to find the same observations are true. Interestingly, while trying to figure out what genes are causative nicotine resistance genes, we discovered a complex 22bp deletion within *Ugt86Dd* that creates a frameshift mutation in the second exon that leads to a premature stop codon and strongly appears to have variable effects on nicotine resistance. Moreover, the complex 22bp is found in natural populations with the deletion event occurring at a frequency ~0.02. This rare allele also appears to be mildly deleterious, supporting the notion that loss of function variants create variable effects for complex traits, specifically xenobiotic resistance, and should be investigated for other compound resistance to determine what roles loss of function variants have on those phenotypes.

In conclusion, results presented in chapter II of this dissertation describe an unbiased screen for lifespan using mated females that revealed metabolic and defense response genes that might contribute to lifespan variation. These data are consistent with what previous studies have found and provides more confidence that these types of genes truly contribute to lifespan variation. Results presented in chapter III provides evidence that *Cyp28d1* and *Cyp28d2* likely harbor segregating variation that affects nicotine resistance, and the complex 22bp deletion in *Ugt86Dd* is likely one of the many causative variants in the DSPR for nicotine resistance. Moreover, we also provide an elegant way to move from QTL to gene via the array of functional assays presented. These data add a significant contribution on of what genes and variants likely contribute to nicotine resistance in natural populations. Overall, both of these studies highlight the complexity of complex traits and the gapping lack of knowledge of the genes that harbor variation for various traits. Thus, strongly suggesting more time and resources need to be poured into teasing apart the genetic basis of complex traits.

Section 4.2 Future Directions

The findings presented in this dissertation provide insights into the complexity of lifespan within *D. melanogaster* and provides insights into what detoxification genes are needed to deal with nicotine. Also, this dissertation presents a way-to get from QTL to gene and some cases the causative variant. However, there are a plethora of unanswered questions. Moving forward, we propose to investigate a few question associated with each of the studies to further our knowledge on the complexity of lifespan and the variable effects of xenobiotic resistance.

Question 1: Are there genotype by subpopulation and genotype by sex QTL within the DSPR?

In chapter II we demonstrated there is significant segregating variation for mated females in subpopulation B of the DSPR. Upon employing our in house QTL mapping software, five significant QTL were detect implicating 11-155 genes per QTL. Underneath several QTL were previously identified lifespan candidates rendering confidence in our mapping data. However, we do not know what the segregating variation for mated females in subpopulation A of the DSPR looks like. Previous studies employing both subpopulations of the DSPR have shown that each subpopulation harbors unique and common QTL for phenotypes within and between subpopulations [22,45,53]. We predict that upon phenotyping subpopulation A of the DSPR, using the same pipeline created for subpopulation B, unique QTL will be uncovered as well as QTL in common with subpopulation B of the DSRP. However, it should be mentioned that each time QTL mapping is employed, a minimal set of QTL are uncovered leaving a plethora of unique QTL and common QTL unknown.

Our study only focused on discovering significant segregating lifespan variation for

mated females in the DSPR and did not examine the variable effects of lifespan that are sex or environmental specific. However, several studies have investigated lifespan in *D. melanogaster* employing the QTL approach to isolate genomic intervals harboring segregating variation for different sexes or environments [8,9,11]. These studies uncovered QTL that were both sex and environmental specific, and could be one of several mechanisms that maintains segregating variation for lifespan in natural populations. Previous support that sex specific effects maintain lifespan variation was demonstrated within and between populations of *D. subobscura* [63]. Thus, it will be of interest to see if sex specific effects in the DSPR maintain segregating variation for lifespan. Thus, we hypothesize QTL mapping in males employing the pipeline utilized in chapter II in both subpopulations of the DSPR will uncover genotype by sex effects.

Question 2: How does the immune system function in creating variable effects on lifespan in *D. melanogaster*?

In chapter II we found that defense response genes are enriched in our RNAseq data sets. Moreover, several other studies have shown that the *Drosophila* immune response influences lifespan and correlated with variation in lifespan [22–28]. While the *Drosophila* immune response is well characterized, it is poorly understood how the immune system creates variable effects on lifespan. Thus, we hypothesize that standing genetic variation that influences lifespan is partly due to variation in the immune response. What has been shown over multiple studies are immune genes in the innate immune system increase as a function of age. Thus, NF-κB, Toll, and IMD signaling pathways are likely important in promoting variation lifespan in the DSPR, but how?

There are several ideas that could make logical sense but need to be investigated. First, the effects we see with immune gene enrichment has no effect on lifespan. This idea is unlikely

when, observing age related growth of *E. coli*, old *Drosophila* compared to young *Drosophila* have trouble suppressing the growth of introduced *E. coli,* suggesting the innate immune system functions less effectively increasing the likelihood of death [64]. However, this does not answer the question, if immune genes harbor variable effects on lifespan. Second, an alternative idea would be upon the introduction of increased pathogen loads, short-lived and long-lived old RILs increase gene expression to deal with the foreign invaders, but the short-lived old RILs have dysregulation, or lose or transcriptional regulation of the immune system from mutations in genes that are responsible for the response. Thus, the short-lived old RIls can not handle increased pathogen loads that results in death. To examine this idea we only need to look at RILs that are classified as short and long-lived RILs and do a variant search in immune genes to uncover variants that do not change gene expression but gene function. This is exactly what we can do with the soon availability of *PacBio*® sequence data. Third, variation in immune gene expression is a direct response to epigenetic alterations, such as histone modifications and chromosome remodeling that come with aging [65]. To examine this idea a simple DNAse sensitivity assay can be performed looking at differential cleavage of open DNA and correlate with areas of the genome that harbor defense response genes.

Question 3: What other Ugt genes have an effect on nicotine resistance under QTL4?

Broadly, we understand how detoxification works as a general process [40], but we have a poor understanding on what genes are utilized to break down specific compounds and an even less understanding what genes harbor variable effects on the detoxification process. Specifically, we do not know what genes or variants have variable effects on nicotine metabolism. There has been some studies that have highlighted that variation in nicotine resistance exist in natural populations [43–45,66] and the transposable element, Accord, within *Cyp6g1* is correlated with

increased nicotine resistance via increasing *Cyp6g1* expression [56]. However, our study [45] and other studies find little continuity in the results. Thus, there is likely a plethora of undiscovered genes and variants that contribute to nicotine resistance.

In chapter III, a bulk of the studies implicated the phase II gene, *Ugt86Dd* as a causative gene that harbors a segregating complex 22bp InDel variant that likely contributes to variation in nicotine resistance. However, a more nuance part of this study demonstrated our QCT data for QTL4 likely harbors multiple genes that contribute to variation in nicotine resistance. Hence, why we examined two other phase II genes, *Ugt86Dh* and *Ugt86Dj*, which had insertional mutational lines available but yielded non-significant results. It should be noted that *Ugt86Dj* was marginally non-significant, and with a higher sample size might reveal a small effect on nicotine resistance. Although, performing QCTs with insertional mutants for every gene under QTL4 seems simple as first, but upon inspection not every gene has insertional mutants available. Furthermore, many insertional mutants available harbor p-elements that are close to or at what appears to be the start/end of the gene. Leaving the possibility these inserts actually do not disrupt the gene's function. Thus, to examine which genes are causative for nicotine resistance, the CRISPR/cas9 system will likely be a better approach to employ to create DSBs that disrupt gene function. Then we can examine the difference in resistance between nonCRISPR and CRISPR lines at each loci. If a difference is observed, then that gene likely contributes to variation in nicotine resistance. This new data would shed light onto more genes and variants that can have variable effects on nicotine resistance and would allow us to start to assemble gene networks to understand the genetic pathways important for nicotine resistance.

Question 4: What role does *Cyp28d1* and *Cyp28d2* play in creating variation in nicotine resistance?

In chapter III the data demonstrates strong evidence that the phase I gene, *Cyp28d1,* is functionally important for nicotine resistance ubiquitously and within the anterior midgut, and that segregating variation exist at this locus between A3 and A4 of the DSPR. However, the data collected for the phase I gene, *Cyp28d2,* is not as strong as for *Cyp28d1*. The data establishes a functional role for nicotine resistance ubiquitously, and that segregating variation affecting nicotine resistance exists at this locus between A3 and A4 of the DSPR. Nevertheless, we still do not have a great understanding what tissues these genes have a role in, and we do not have any candidate variants in or around the genes that would point to why segregating variation exist at these loci in the DSPR. To address what tissues these genes have a functional role in, more tissue specific RNAi along with tissue specific over-expression assays need to be performed. Likely, tissues of the digestive tract and nervous system will be a prime target. There have been several studies showing nicotine has a cascade of effects in the nervous system [67–69] and our data suggest the digestive tract is involved in nicotine metabolism. To address what variants in the genes are causative, Sanger sequence both A3's and A4's version *Cyp28d1* and *Cyp28d2* to confirm no InDels where missed as the case for *Ugt86Dd*. If any InDels were missed, proceed to verify the causal effects of the InDel. However, if no InDels are uncovered, then employ the CRISPR/cas9 system with ssODN that contains the gene and variable lengths of the intergenic region outside the gene can be employed. In this approach we would swap the A3 version of the gene and the variable intergenic regions into A4 and compare to A4 for diminished nicotine resistance. Furthermore, swapping the A4 version of the gene and the variable intergenic regions into A3 would need to be performed to rescue resistance. Any differences in resistance observed would find the causative variant(s) or small set of variants that harbors the true causative variant(s).

Literature Cited

1. Chow CY, Kelsey KJP, Wolfner MF, Clark AG. Candidate genetic modifiers of retinitis pigmentosa identified by exploiting natural variation in Drosophila. Hum. Mol. Genet. 2016;25:651–9.

2. Chow CY. Bringing genetic background into focus. Nat. Rev. Genet. 2016;17:63–4.

3. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. Nat. Rev. Genet. 2006;7:436–48.

4. Reynolds RM, Phillips PC. Natural Variation for Lifespan and Stress Response in the Nematode Caenorhabditis remanei. PLOS ONE. 2013;8:e58212.

5. Ivanov DK, Escott-Price V, Ziehm M, Magwire MM, Mackay TFC, Partridge L, et al. Longevity GWAS Using the Drosophila Genetic Reference Panel. J. Gerontol. A. Biol. Sci. Med. Sci. 2015;

6. Pickering AM, Lehr M, Miller RA. Lifespan of mice and primates correlates with immunoproteasome expression. J. Clin. Invest. 2015;125:2059–68.

7. Nuzhdin SV, Khazaeli AA, Curtsinger JW. Survival Analysis of Life Span Quantitative Trait Loci in Drosophila melanogaster. Genetics. 2005;170:719–31.

8. Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF. Sex-specific quantitative trait loci affecting longevity in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 1997;94:9734–9.

9. Leips J, Mackay TFC. Quantitative Trait Loci for Life Span in Drosophila melanogaster: Interactions With Genetic Background and Larval Density. Genetics. 2000;155:1773–88.

10. Leips J, Mackay TFC. The Complex Genetic Architecture of Drosophila Life Span. Exp. Aging Res. 2002;28:361–90.

11. Vieira C, Pasyukova EG, Zeng Z-B, Hackett JB, Lyman RF, Mackay TFC. Genotype-environment interaction for quantitative trait loci affecting life span in Drosophila melanogaster. Genetics. 2000;154:213–27.

12. Curtsinger JW, Khazaeli AA. Lifespan, QTLs, age-specificity, and pleiotropy in Drosophila. Mech. Ageing Dev. 2002;123:81–93.

13. Geiger-Thornsberry GL, Mackay TFC. Quantitative trait loci affecting natural variation in Drosophila longevity. Mech. Ageing Dev. 2004;125:179–89.

14. Reiwitch SG, Nuzhdin SV. Quantitative trait loci for lifespan of mated Drosophila melanogaster affect both sexes. Genet Res. 2002;80:225–30.

15. Wang M-H, Lazebny O, Harshman LG, Nuzhdin SV. Environment-dependent survival of Drosophila melanogaster: a quantitative genetic analysis. Aging Cell. 2004;3:133–40.

16. Wilson RH, Morgan TJ, Mackay TFC. High-Resolution Mapping of Quantitative Trait Loci Affecting Increased Life Span in Drosophila melanogaster. Genetics. 2006;173:1455–63.

17. Lai C-Q, Parnell LD, Lyman RF, Ordovas JM, Mackay TFC. Candidate genes affecting Drosophila life span identified by integrating microarray gene expression analysis and QTL mapping. Mech. Ageing Dev. 2007;128:237–49.

18. Forbes SN, Valenzuela RK, Keim P, Service PM. Quantitative trait loci affecting life span in replicated populations of Drosophila melanogaster. I. Composite interval mapping. Genetics. 2004;168:301–11.

19. Defays R, Gomez FH, Sambucetti P, Scannapieco AC, Loeschcke V, Norry FM. Quantitative trait loci for longevity in heat-stressed Drosophila melanogaster. Exp. Gerontol. 2011;46:819–26.

20. King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, et al. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 2012;22:1558–66.

21. King EG, Macdonald SJ, Long AD. Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics. 2012;191:935–49.

22. Burke MK, King EG, Shahrestani P, Rose MR, Long AD. Genome-wide association study of extreme longevity in Drosophila melanogaster. Genome Biol. Evol. 2013;evt180.

23. Carlson KA, Gardner K, Pashaj A, Carlson DJ, Yu F, Eudy JD, et al. Genome-Wide Gene Expression in relation to Age in Large Laboratory Cohorts of Drosophila melanogaster. Genet. Res. Int. 2015;2015:835624.

24. De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. Proc. Natl. Acad. Sci. U. S. A. 2001;98:12590–5.

25. Garigan D, Hsu A-L, Fraser AG, Kamath RS, Ahringer J, Kenyon C. Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. Genetics. 2002;161:1101–12.

26. Garsin DA, Villanueva JM, Begun J, Kim DH, Sifri CD, Calderwood SB, et al. Long-lived C. elegans daf-2 mutants are resistant to bacterial pathogens. Science. 2003;300:1921.

27. Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, et al. Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. 2004;101:7663–8.

28. Murphy CT, McCarroll SA, Bargmann CI, Fraser A, Kamath RS, Ahringer J, et al. Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. Nature. 2003;424:277–83.

29. Pletcher SD, Macdonald SJ, Marguerie R, Certa U, Stearns SC, Goldstein DB, et al. Genome-wide transcript profiles in aging and calorically restricted Drosophila melanogaster. Curr Biol. 2002;12:712–23.

30. Felix TM, Hughes KA, Stone EA, Drnevich JM, Leips J. Age-Specific Variation in Immune Response in Drosophila melanogaster Has a Genetic Basis. Genetics. 2012;191:989–1002.

31. Ren C, Webster P, Finkel SE, Tower J. Increased internal and external bacterial load during Drosophila aging without life-span trade-off. Cell Metab. 2007;6:144–52.

32. Zerofsky M, Harel E, Silverman N, Tatar M. Aging of the innate immune response in Drosophila melanogaster. Aging Cell. 2005;4:103–8.

33. DeVeale B, Brummel T, Seroude L. Immunity and aging: the enemy within? Aging Cell. 2004;3:195–208.

34. Libert S, Chao Y, Chu X, Pletcher SD. Trade-offs between longevity and pathogen resistance in Drosophila melanogaster are mediated by NFkappaB signaling. Aging Cell. 2006;5:533–43.

35. Myllymäki H, Rämet M. JAK/STAT Pathway in Drosophila Immunity. Scand. J. Immunol. 2014;79:377–85.

36. Dushay MS, Asling B, Hultmark D. Origins of immunity: Relish, a compound Rel-like gene in the antibacterial defense of Drosophila. Proc. Natl. Acad. Sci. U. S. A. 1996;93:10343–7.

37. Woodcock KJ, Kierdorf K, Pouchelon CA, Vivancos V, Dionne MS, Geissmann F. Macrophage-Derived upd3 Cytokine Causes Impaired Glucose Homeostasis and Reduced Lifespan in Drosophila Fed a Lipid-Rich Diet. Immunity. 2015;42:133–44.

38. Brown S, Hu N, Hombría JC. Identification of the first invertebrate interleukin JAK/STAT receptor, the Drosophila gene domeless. Curr. Biol. CB. 2001;11:1700–5.

39. About Zika Virus Disease | Zika virus | CDC [Internet]. [cited 2016 Jun 5]. Available from: http://www.cdc.gov/zika/about/index.html

40. Xu C, Yong-Tao C, Kong A-NT. Induction of phase I, II and III drug metabolism/transport by xenobiotics. Arch Pharm Res. 2005;28:249–68.

41. Ffrench-Constant RH. The Molecular Genetics of Insecticide Resistance. Genetics. 2013;194:807–15.

42. Hill R, Wyse G, Anderson M. Nervous System Organization and Biological Clocks. Anim. Physiol. 3rd ed. Sinauer Associates, Inc.; 2012.

43. Carrillo R, Gibson G. Unusual genetic architecture of natural variation affecting drug resistance in Drosophila melanogaster. Genet Res. 2002;80:205–13.

44. Passador-Gurgel G, Hsieh W-P, Hunt P, Deighton N, Gibson G. Quantitative trait transcripts for nicotine resistance in Drosophila melanogaster. Nat Genet. 2007;39:264–8.

45. Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. Genetics. 2014;198:45–57.

46. Tan Q, Zhao JH, Zhang D, Kruse TA, Christensen K. Power for genetic association study of human longevity using the case-control design. Am. J. Epidemiol. 2008;168:890–6.

47. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. 2001;69:124–37.

48. Mackay TFC. Quantitative Trait Loci In Drosophila. Nat. Rev. Genet. 2001;2:11–20.

49. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, et al. A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 2009;5:e1000551.

50. Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS, et al. Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Res. 2011;21:1213–22.

51. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. 2004;36:1133–7.

52. McNeil CL, Bain CL, Macdonald SJ. Multiple Quantitative Trait Loci Influence the Shape of a Male-Specific Genital Structure in Drosophila melanogaster. G3 Bethesda. 2011;1:343–51.

53. Najarro MA, Hackett JL, Smith BR, Highfill CA, King EG, Long AD, et al. Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. PLoS Genet. 2015;11:e1005663.

54. Kuehl GE, Murphy SE. N-glucuronidation of nicotine and cotinine by human liver microsomes and heterologously expressed UDP-glucuronosyltransferases. Drug Metab. Dispos. Biol. Fate Chem. 2003;31:1361–8.

55. Sakamoto M, Itoh T, Tukey RH, Fujiwara R. Nicotine regulates the expression of UDP-glucuronosyltransferase (UGT) in humanized UGT1 mouse brain. Drug Metab. Pharmacokinet. 2015;30:269–75.

56. Li X, Bai S, Cass BN. Accord insertion in the 5' flanking region of CYP6G1 confers nicotine resistance in Drosophila melanogaster. Gene. 2012;502:1–8.

57. O'Sullivan BP, Freedman SD. Cystic fibrosis. Lancet Lond. Engl. 2009;373:1891–904.

58. Highsmith WE, Burch LH, Zhou Z, Olsen JC, Boat TE, Spock A, et al. A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. N. Engl. J. Med. 1994;331:974–80.

59. Cuppens H, Lin W, Jaspers M, Costes B, Teng H, Vankeerberghen A, et al. Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. J. Clin. Invest. 1998;101:487–96.

60. MacArthur DG, Tyler-Smith C. Loss-of-function variants in the genomes of healthy humans. Hum. Mol. Genet. 2010;19:R125–30.

61. Gaedigk A, Blum M, Gaedigk R, Eichelbaum M, Meyer UA. Deletion of the entire cytochrome P450 CYP2D6 gene as a cause of impaired drug metabolism in poor metabolizers of the debrisoquine/sparteine polymorphism. Am. J. Hum. Genet. 1991;48:943–50.

62. De Morais SM, Wilkinson GR, Blaisdell J, Nakamura K, Meyer UA, Goldstein JA. The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. J. Biol. Chem. 1994;269:15419–22.

63. Smith JM. The Effects of Temperature and of Egg-Laying on the Longevity of Drosophila subobscura. Sci. Aging Knowl. Environ. 1958;2002:cp15.

64. Kim YS, Nam HJ, Chung HY, Kim ND, Ryu JH, Lee WJ, et al. Role of xanthine dehydrogenase and aging on the innate immune response of Drosophila. J. Am. Aging Assoc. 2001;24:187–93.

65. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. Cell. 2013;153:1194–217.

66. Hall LM, Von Borstel RW, Osmond BC, Hoeltzli SD, Hudson TH. Genetic variants in an acetylcholine receptor from Drosophila melanogaster. FEBS Lett. 1978;95:243–8.

67. Baiton RJ, Tsai LTY, Singh CM, Moore MS, Neckameyer WS, Heberlein U. Dopamine modulates acute responses to cocaine, nicotine and ethanol in Drosophila. Curr. Biol. 2000;10:187–94.

68. Kaun KR, Devineni AV, Heberlein U. Drosophila melanogaster as a model to study drug addiction. Hum Genet. 2012;131:959–75.

69. Wolf FW, Heberlein U. Invertebrate Models of Drug Abuse. J Neurobiol. 2003;54:161–78.