Individual variation in alcohol consumption in human populations is determined by genetic, environmental, social and cultural factors. In contrast to humans, genetic contributions to complex behavioral phenotypes can be readily dissected in Drosophila, where both the genetic background and environment can be controlled and behaviors quantified through simple high-throughput assays. Here, we measured voluntary consumption of ethanol in ~3000 individuals of each sex from an advanced intercross population derived from 37 lines of the Drosophila melanogaster Genetic Reference Panel. Extreme quantitative trait loci mapping identified 385 differentially segregating allelic variants located in or near 291 genes at $P < 10^{-8}$. The effects of single nucleotide polymorphisms associated with voluntary ethanol consumption are sex-specific, as found for other alcohol-related phenotypes. To assess causality, we used RNA interference knockdown or $P\{\text{MiET1}\}$ mutants and their corresponding controls and functionally validated 86% of candidate genes in at least one sex. We constructed a genetic network comprised of 23 genes along with a separate trio and a pair of connected genes. Gene ontology analyses showed enrichment of developmental genes, including development of the nervous system. Furthermore, a network of human orthologs showed enrichment for signal transduction processes, protein metabolism and developmental processes, including nervous system development. Our results show that the genetic architecture that underlies variation in voluntary ethanol consumption is sexually dimorphic and partially overlaps with genetic factors that control variation in feeding behavior and alcohol sensitivity. This integrative genetic architecture is rooted in evolutionarily conserved features that can be extrapolated to human genetic interaction networks.

Throughout history, alcohol consumption has been integral to human culture, either in ceremonial or social contexts. Whereas social drinking is an acceptable norm in societies where alcohol consumption is allowed, excessive alcohol intake is a risk factor for alcohol abuse and alcoholism. Individuals, however, vary significantly in their propensity toward alcohol consumption. Several genetic studies in human populations have implicated genes associated with alcohol-related phenotypes (Bierut et al. 2010, 2012; Edenberg et al. 2010; Frank et al. 2012; Gelernter et al. 2014; Heath et al. 2011; Kendler et al. 2011; Park et al. 2013; Treutlein et al. 2008; Treutlein & Rietschel 2011). The diverse spectrum of alcohol-related phenotypes poses the question whether different elements of the genome determine the manifestation of each phenotype or whether all alcohol-related phenotypes arise from a common genetic architecture. It is challenging to address this question in human populations due to lack of control of genetic backgrounds and confounding environmental factors affecting alcohol-related phenotypes, such as stress and neuropsychiatric disorders, as well as developmental history and the social environment (Clarke et al. 2016; Edenberg & Foroud 2013; Palmer et al. 2012, 2015; Rietschel & Treutlein 2013).

Drosophila melanogaster is a powerful model system that enables comprehensive genome-wide genetic analyses to identify genes and genetic networks associated with different aspects of alcohol sensitivity, as both the genetic background and environmental rearing conditions can be controlled precisely and ethanol intake can be quantified accurately. The D. melanogaster Genetic Reference Panel (DGRP), a population of 205 sequenced and well-annotated inbred wild-derived lines (Huang et al. 2014; Mackay et al. 2012), has previously enabled genome-wide association (GWA) analyses of alcohol sensitivity to identify genetic networks associated with variation in ethanol knockdown time (Morozova et al. 2015). These networks served as contextual blueprints for orthologous networks of corresponding human genes (Bier 2005; Hu et al. 2011; Lloyd & Taylor 2010).

Here, we ask to what extent variation in voluntary alcohol consumption has a shared genetic basis with previously documented variation in the inebriating effects of ethanol.

Keywords: Advanced intercross, behavioral genetics, CAFE assay, Drosophila melanogaster Genetic Reference Panel, extreme QTL mapping, feeding behavior, functional genomics, mutational analysis, network, RNAi, single nucleotide polymorphism, translational genomics

Received 26 January 2017, revised 24 May 2017, accepted for publication 14 June 2017

Individual variation in alcohol consumption in human populations is determined by genetic, environmental, social and cultural factors. In contrast to humans, genetic contributions to complex behavioral phenotypes can be readily dissected in Drosophila, where both the genetic background and environment can be controlled and behaviors quantified through simple high-throughput assays. Here, we measured voluntary consumption of ethanol in ~3000 individuals of each sex from an advanced intercross population derived from 37 lines of the Drosophila melanogaster Genetic Reference Panel. Extreme quantitative trait loci mapping identified 385 differentially segregating allelic variants located in or near 291 genes at $P < 10^{-8}$. The effects of single nucleotide polymorphisms associated with voluntary ethanol consumption are sex-specific, as found for other alcohol-related phenotypes. To assess causality, we used RNA interference knockdown or $P\{\text{MiET1}\}$ mutants and their corresponding controls and functionally validated 86% of candidate genes in at least one sex. We constructed a genetic network comprised of 23 genes along with a separate trio and a pair of connected genes. Gene ontology analyses showed enrichment of developmental genes, including development of the nervous system. Furthermore, a network of human orthologs showed enrichment for signal transduction processes, protein metabolism and developmental processes, including nervous system development. Our results show that the genetic architecture that underlies variation in voluntary ethanol consumption is sexually dimorphic and partially overlaps with genetic factors that control variation in feeding behavior and alcohol sensitivity. This integrative genetic architecture is rooted in evolutionarily conserved features that can be extrapolated to human genetic interaction networks.
(Morozova et al. 2015). To uncover the genetic underpinnings that underlie variation in voluntary alcohol consumption, we measured intake of an ethanol-supplemented sucrose solution by individual flies of an advanced intercross population (AIP), derived from 37 DGRP lines with maximal homozygosity, minimal relatedness and absence of chromosomal inversions and the endosymbiont Wolbachia (Garlapow et al. 2016). This AIP was generated through a round-robin crossing design, and was maintained subsequently by random mating and a large population size. The advantages of the AIP compared to a GWA study across the DGRP are: (1) we gain statistical power, because we are generating a vast number of unique genotypes as a result of recombination, as compared to only 205 genotypes represented by the DGRP; (2) because the AIP is started from a round-robin crossing design of the original 37 lines, low-frequency alleles (<5% in the DGRP), which cannot be analyzed in the DGRP because of spurious linkage disequilibrium, but may have large phenotypic effects, are well represented in the starting population. Thus, the AIP allows us to survey the entire allelic spectrum and identify differentially segregating alleles between the extremes of the phenotypic distribution. This ‘extreme quantitative trait loci (QTL)’ mapping strategy (Ehrenreich et al. 2010; Huang et al. 2012) enabled us to uncover genes and networks associated with variation in ethanol consumption in Drosophila, compare this network with previous observations on the genetic underpinnings of alcohol inebriation, and translate this network into a network of human orthologs.

Materials and methods

Drosophila stocks

We used an AIP derived from 37 DGRP lines (DGRP_41, DGRP_42, DGRP_45, DGRP_59, DGRP_83, DGRP_91, DGRP_129, DGRP_158, DGRP_177, DGRP_195, DGRP_208, DGRP_217, DGRP_228, DGRP_229, DGRP_235, DGRP_239, DGRP_307, DGRP_315, DGRP_367, DGRP_371, DGRP_375, DGRP_379, DGRP_391, DGRP_399, DGRP_427, DGRP_439, DGRP_491, DGRP_508, DGRP_509, DGRP_517, DGRP_703, DGRP_757, DGRP_765, DGRP_799, DGRP_808, DGRP_843 and DGRP_900), which are free of inversions, free of Wolbachia infection, are maximally homozygous and maximally unrelated. We crossed these lines in a round-robin design as described previously (Carbone et al. 2016; Huang et al. 2012). We reared flies in bottles on cornmeal/molasses/agar medium under standard culture conditions (25°C, 12:12 h light/dark cycle). To maintain the population and minimize genetic drift, we combined at a constant population size of 800. To minimize natural selection and 40 females in 10 new bottles to generate the next population each generation offspring from 10 bottles and distributed 40 males and 40 females in 10 new bottles to generate the next population at a constant population size of 800. To minimize natural selection through larval competition, egg laying was restricted to 24 h. The experiments described here began at G14.

We obtained 25 RNA interference (RNAi) transgenic fly strains of the phiC31 (KK) RNAi library (AAdamtS-A, alpha-Man-lb, CG12506, CG14137, CG18418, CG31638, CG34362, CG45263, CG6231, CG6660, CG7368, CG7514, cpo, D19A, Dhc88D, fred, grass, htt, klg, Ian, Men-b, Hmp2, mso, Myo31DFD and sinal), three strains of the P-element (GD) RNAi library (daily, CG17097 and loco) together with the corresponding progenitor lines (60016 and 60020, respectively) from the Vienna Drosophila RNAi Center (VDRC) (Dietzl et al. 2007). These RNAi lines are predicted not to have off-target effects. These lines and the appropriate progenitor controls were crossed to a weak Ubiquitin-GAL4 driver (Garlapow et al. 2015). Eight MiETTI mutants (CG12910, CG33158, CG45186, dptr, rg, Slc45-1, sr and Trim9) and their co-isogenic control w1118; 2 iso 3 iso Bellen et al. 2011) were obtained from the Bloomington Drosophila stock center (Bloomington, IN, USA).

Quantitative RT-PCR

We quantified mRNA levels of RNAi knockdown lines by quantitative reverse transcriptase-polymerase chain reaction (RT-PCR) with the SYBR Green detection method (Maxima SYBR Green; Thermo Scientific), as described previously, with two biological and two technical replicates per line per sex using glyceraldehyde-3-phosphate dehydrogenase (Gpdh) as the internal standard (Morozova et al. 2015). Briefly, total RNA was extracted using Trizol™ Reagent (Invitrogen™, Thermo Fisher Scientific, Waltham, MA, USA) and cDNA was generated from 500 ng of total RNA by reverse transcription using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA). Primers were designed to span exon–intron junctions using Primer3Plus v2.4.0 (Untergasser et al. 2007). Expression of each gene in each biological replicate for each RNAi and control line and sex was normalized relative to the appropriate Gpdh expression level using ΔCt values (Morozova et al. 2015). Statistically significant differences in gene expression levels between RNAi knockdown and control lines were determined by Student’s t-tests on ΔCt values.

Ethanol consumption

To measure consumption of ethanol-supplemented food, we used a modified Capillary Feeding (CAFE) assay (Garlapow et al. 2016; Ja et al. 2007). At least 24 h prior to the assay, we placed individual 3–7-day-old virgin flies in vials containing culture medium following CO2 anesthesia. To measure consumption of ethanol-supplemented food, we transferred each fly into individual vials containing 3 ml of 1.5% agar medium and one 5-μl capillary tube (71900-5, Kimble Chase Life Science, Rockwood, TN, USA) inserted through a foam plug. The capillary contained a 4% ethanol (v/v) in 8% sucrose solution with mineral oil on the top (Fig. 1a). We placed 130 of the CAFE vials in a humidified chamber (80% humidity) that contained 20 vials of 10 ml water. To account for evaporation, we placed 10 vials with capillary tubes containing 4% ethanol in 8% sucrose without flies in the same humidified chamber. We marked the level of fluid in each capillary tube at the beginning of the assay. Flies were allowed to feed for 24 h, after which we measured the level of fluid in each capillary. Total food consumption was calculated as the difference (mm) in fluid levels in the capillaries, corrected for the average evaporation that occurred in the negative control vials, and converted from mm to volume consumed (μl/fly) by dividing by 15 (the capillaries are calibrated such that 75 mm corresponds to 5 μl).

Each day, the top and bottom 10% of consumers for each sex were collected and flash-frozen along with the same number of randomly selected flies. Samples of 300 high consuming and low consuming flies of each sex were accumulated over time and pooled into three replicates of 100 flies each for bulk DNA sequencing. We also sequenced two replicates of 100 randomly selected males and females. In total, we scored ~3000 males and 3000 females for ethanol-supplemented sucrose consumption (hereafter referred to as ‘ethanol consumption’).

DNA sequencing

We homogenized whole flies using a Tissuelyser (Qiagen, Inc., Germantown, MD, USA) and extracted genomic DNA using the Gentra Puregene Tissue kit according to the manufacturer’s protocol (Qiagen Sciences, MD, USA). Genomic DNA was fragmented to 300–400 bp using a Covaris S220 sonicator (Covaris, Inc, Woburn, MA, USA). We used 300 ng of fragmented DNA to produce barcoded DNA libraries using NEXTflex™ DNA Barcodes (Bioo Scientific, Inc., Austin, TX, USA) according to an Illumina TrueSeq compatible protocol (Illumina, Inc., San Diego, CA, USA). Libraries were quantified using Qubit dsDNA HS Kits (Life Technologies, Inc., Eugene, OR, USA) and a Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA) to calculate molarity. Libraries were then diluted to equal molarity and re-quantified, and all 16 barcoded samples were pooled. Pooled library samples were quantified again to calculate final molarity and then denatured and diluted to 16 μM. They were clustered on an Illumina
Genetics of alcohol consumption in *Drosophila*

We performed an extreme QTL mapping analysis (Ehrenreich et al. 2012) to identify single nucleotide polymorphisms (SNPs) with significant changes in allele frequency between the pools of high, control or low consumers. Barcoded paired-end sequence reads were demultiplexed using the Illumina pipeline v1.9, and aligned to the *D. melanogaster* reference genome (BDG5) using Burrows-Wheeler Aligner (BWA v0.7.12) (Li & Durbin 2010). GATK (version 3.3) (DePristo et al. 2011) was used to locally realign reads around indels, mark PCR duplicates (with Picard tools version 1.128) and recalibrate base qualities. We piled up high quality bases (Q > 13) using SAMTools version 1.2 (Li et al. 2009) to obtain counts of alleles at SNP sites where the parental lines are polymorphic for the alleles.

Finally, we tested for differences between the high (H) and low (L) alcohol consumer pools, high and control (C) pools, low and control pools and high vs. low together with control pools combined using a $\chi^2$ test, where the test statistic was calculated as:

$$\chi^2 = \sum w_i^2 p_i (1 - p_i) \left( \frac{1}{n_i} + \frac{1}{n_j} \right) + \sum w_j^2 p_j (1 - p_j) \left( \frac{1}{n_i} + \frac{1}{n_j} \right),$$

where $p_i$ and $p_j$ were the weighted (by sequence depth) average allele frequencies in the groups (e.g. high vs. low consumer pools) being compared; $w_i$ and $w_j$ were the weights of replicate pools within each group; $p_i$ and $p_j$ were the allele frequencies in replicate pools in the two groups respectively and $i$ and $j$ are indexed replicates; $n_i$ and $n_j$ were the number of chromosomes; and $d_i$ and $d_j$ were the sequence depths. $P$-values were obtained by finding the probability high with values from the $\chi^2$ distribution with one degree of freedom. We tested 2,336,782 SNPs in the analyses. The Bonferroni-corrected $P$-value for all analyses is $P < 2.14 \times 10^{-8}$. At this stringent $P$-value, the deep sequence coverage, the large numbers of flies sequenced and the number of replicates, we have at least 50% power to detect an allele frequency difference between 0.2 and 0.25 for a range (0.2–0.4) of allele frequency in one of the pools (Fig. S1, Supporting information).

**Functional analysis of candidate genes**

We selected 36 genes with $P$-values exceeding the Bonferroni-corrected threshold for association with ethanol consumption for all pairwise comparisons (high vs. low, high vs. control, high vs. low + control and low vs. control). We functionally tested 36 genes with available *PMET1* mutant lines or RNAi knockdown alleles using the CAFE assay described above. For these experiments, we assessed total consumption for 15 replicates of four single sex flies for each sex and genotype with two capillaries per vial. We measured both the amount of 8% sucrose alone and the amount of 4% ethanol in 8% sucrose consumed for each mutant line and its control for each sex separately.

We assessed whether differences in ethanol intake vs. intake of sucrose alone were significant for each mutant line, separately for males and females, using a mixed model analysis of variance of form $Y = \mu + F + G + F \times G + d + \varepsilon$, where $Y$ is the observed value, $\mu$ is the mean, $F$ indicates the fixed effect of food (ethanol vs. sucrose alone), $G$ indicates mutant or control genotypes (fixed) and $\varepsilon$ is the residual variance. Significance of the $F \times G$ interaction term indicates an effect of the mutation on the level of ethanol consumption.

We also assessed differences in residual variance between mutant alleles of candidate genes and the appropriate controls with pairwise Levene’s tests, separately for ethanol and sucrose consumption, and for males and females (Morgante et al. 2015). Statistical tests were performed using SAS (9.3) software.

**Bioinformatics analysis**

We annotated DNA variants using the gene models in FlyBase release r5.57 (McQuilton et al. 2012). We mapped 291 candidate genes significant at $P < 10^{-5}$ to the physical and genetic interaction databases downloaded from FlyBase release r5.57. We then extracted subnetworks from the global networks whose edges were either a direct connection between candidate genes or bridged by only one gene not among the candidate gene list. We evaluated the significance of the size of the largest cluster among the subnetworks by a randomization test in which we randomly extracted subnetworks with the same number of input genes. The $P$-value was determined by dividing the number of instances where the size of the largest cluster exceeds the
observed largest size by the total number of randomizations ($\alpha = 0.05$) (Antonov et al. 2008). We performed gene ontology enrichment analysis of 23 connected candidate genes using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Dennis et al. 2003; Huang da et al. 2009). Human orthologs were obtained using the DRSC Integrative Ortholog Prediction Tool (DIOPT, version 5.4; http://www.flyrnai.org/diopit; Hu et al. 2011). A gene interaction network for human orthologs was constructed using R-Spider (http://www.bioprofiling.de; Antonov 2011).

**Results**

**Variation in ethanol consumption**

To assess variation in ethanol consumption, we tested 3000 individual males and 3000 individual females of the AIP using the CAFE assay (Fig. 1a). The phenotypic distributions of ethanol consumption were skewed and similar for both sexes, ranging from 0 to 3 \( \mu l / g \) (Fig. 1b,c). We collected three replicates of 100 flies each of the 10% lowest and the 10% highest consumers, along with two replicates of 100 random control flies, for extreme QTL mapping analysis. Most low drinkers did not consume any ethanol, while control flies and high drinkers consumed on average 0.35 and 1.2 \( \mu l \) of ethanol solution, respectively. Thus, we find substantial variation in ethanol consumption among the outbred flies.

**Extreme QTL mapping**

Next, we sought to identify differentially segregating alleles between the control, low and high ethanol consumption groups, separately for males and females. We sequenced DNA from the 16 pools of flies and performed multiple comparisons between the selected and control groups, separately for males and females (Table S1). For each comparison, we considered variants with a Bonferroni-corrected \( P \) value of \( P < 2.14 \times 10^{-6} \) to be significant (Fig. 2). In females (males), we identified 12 (14) SNPs located in or near 7 (10) genes with significant differences in allele frequency between low and high consumers (high vs. low); and 29 (27) SNPs in or near 23 (22) genes with significant differences in allele frequency between high consumers and the control group (high vs. control). As the control group was biased toward low consumption (Fig. 1), we also compared high consumers with low and control consumers combined and identified 24 (62) SNPs located in or near 15 (61) genes with significant differences in allele frequency in females (males). Finally, we identified 13 (17) SNPs located in or near 10 (10) genes associated with significant allele frequency differences between low consumers and the control group (low vs. control) in females (males) (Fig. 2; Table S1). These 10 genes were different for males and females.

Summed across all comparisons, we mapped 71 and 100 SNPs with significant differences in allele frequencies in females and males, respectively. However, only one intergenic SNP (3R_21930845) was common between males and females, and no genes were common to both sexes. We did find some overlap at the SNP and gene levels between the different comparisons within sex groups (Table S1). For example, 3R_15442665_SNPs in CG6231 is common between the low vs. control, high vs. low + control and high vs. low comparisons in males. Similarly, 3L_61826262_SNPs, which is located upstream of D19A and CG7368, is common between the high vs. control and low vs. control comparisons in females. Three SNPs (3L_16452334_SNPs, 3L_16452332_SNPs and 3L_16452331_SNPs) located in CG33158, and 3L_11591892_SNPs located in CG7368 and downstream of CG14137 were in common between the high vs. low and high vs. low + control comparisons in females.

CG6231 is involved in transmembrane transport (Tweedie et al. 2009), CG33158 is associated with GTPase activity and CG7368 is involved in phagocytosis and transcription (Stroschein-Stevenson et al. 2006).

At the gene level, we also identified Dhc98D, CadN, CG13506, dally, jing, CG18418 and CG7514, CG6660, klg, cpo, cnc, drpr, loco, lsn and nAChRalpha6 in multiple comparisons.

Dhc98D is involved in ATPase activity (Tweedie et al. 2009), the gene product of CadN is involved in cell adhesion (Yonekura et al. 2007) and axon guidance (Berger et al. 2008a) and dally is involved in axonogenesis (Grueber et al. 2007) and sensory organ development (Kirkpatrick et al. 2006). CG18418 and CG7514 are associated with malate and mitochondrial transport, and klg has been associated with behavioral response to ethanol (Berger et al. 2008b; Morozova et al. 2015).

Thus, our extreme QTL mapping analyses identified many candidate SNPs and genes plausibly associated with variation in ethanol consumption, indicating the polygenic nature of this trait in Drosophila.

**Functional analysis of candidate genes**

The SNPs with significant changes in frequencies between high and low, high and control and low and control DNA pools are either causal SNPs or in linkage disequilibrium with causal alleles affecting ethanol consumption. Some may also represent false positives resulting from genetic drift. Furthermore, as we only measured individual flies in the AIP for consumption of ethanol, we could not disentangle ethanol consumption from sucrose consumption, which is also genetically variable in the DGRP (Garlapow et al. 2016).

To address these issues, we conducted functional analyses using either RNAi knockdown of gene expression or \( P(MET1) \) insertion mutants corresponding to candidate genes that harbor SNPs associated with differences in ethanol consumption. We measured consumption of ethanol and sucrose alone for 36 mutants and their corresponding controls to separate the effects of sucrose consumption from ethanol consumption. We chose genes for these functional analyses for which associated SNPs passed the Bonferroni-corrected significance threshold and have human orthologs (Table S2). These genes are involved in a wide range of biological processes, including metabolic processes (i.e., CG17097, Men-b, msn and sr) (Ross et al. 2003; Tweedie et al. 2009; Ugrankar et al. 2015); proteolysis (i.e., AdamTS-A, CG42750 and grass) (Tweedie et al. 2009); and nervous system development (i.e., CG7368, dally, htt, rg, sr, and Trim9) (de la Pompa et al. 1989; Grueber et al. 2007; Neumuller et al. 2011; Sepp et al. 2008; Song et al. 2011; Volders et al. 2012). Mmp2 has also been implicated previously to affect natural variation in sucrose consumption (Garlapow et al. 2016).
Genetics of alcohol consumption in *Drosophila*

Figure 2: Extreme QTL mapping analysis. Manhattan plots show −log10-transformed P-values for associations of SNPs with differences in ethanol consumption between high and low, high and control, high and low + control and low vs. control samples, separately for females and males. The X-axes indicate chromosomal locations. The dashed lines correspond to the Bonferroni-corrected significance level (P = 2.14 x 10^-8). Dark red and dark blue dots indicate SNPs with significant differences in allele frequencies averaged across replicates at or above the Bonferroni-corrected threshold in females and males, respectively.

To assess the extent of knockdown of the target gene by RNAi, we performed quantitative RT-PCR on a sample of 15 RNAi mutants and control in males and females separately (Fig. S2). With the weak ubiquitin driver used in these experiments, we observed extensive variation among the extent of knockdown of the target gene ranging from 10% to 95% with average knockdown of 34% in males and 45% in females, similar to a previous report on variation of RNAi knockdown of odorant binding protein gene expression (Swarup et al. 2011).

Across both sexes, 64% of the mutants tested showed a significant difference in ethanol intake from the control (Fig. 3; Table S3). Moreover, 86% (31 out of 36 mutants) showed a significant difference in ethanol intake from the control in at least one sex (Fig. 3). Most mutants showed a reduction in ethanol intake compared to control, but some showed increased ethanol consumption, e.g. AdamTS-A, CG12910, CG33158, dpr, Slc45-1 and Trim9 for females, and CG17097, CG45186, rg and Trim9 for males. In 14 cases, we found statistically significant effects in the same direction for males and females. However, CG17097, a gene with a transcript predicted to have lipase activity (Mueller et al. 2004), showed opposite effects between the sexes. Increased ethanol consumption in the mutant background indicates that the wild-type allele limits the amount of alcohol consumed. In contrast, low consumption of alcohol in the mutant background indicates that the wild-type allele promotes intake of alcohol.
Figure 3: Functional analyses of candidate genes. Ethanol consumption compared to sucrose consumption alone of 36 candidate female (a) and male (b) mutants. Data are presented as the difference between ethanol and sucrose consumption ± SEM. The color bar indicates the significance levels for both (a) and (b). Deviations below zero indicate less ethanol consumption and deviations above zero indicate increased ethanol consumption relative to sucrose consumption.

There was no correlation between the extent of RNAi knockdown and phenotypic effect. Failure to confirm some of the genes that harbor alleles with differential allele frequencies could be due to false positives or SNP location. For example, *klg* was previously associated with sucrose consumption (Garlapow et al. 2016) and ethanol sensitivity (Berger et al. 2008b; Morozova et al. 2015), and harbors 12 SNPs that vary across the 37 founder DGRP lines, but only one SNP (3R_18749072_SNPs) showed significant association with ethanol consumption. This SNP is located in both a *klg* intron and upstream of the promotor of *CG6660*. Whereas *klg* was not functionally validated, *CG6660* was confirmed by mutational analysis in both sexes (Tables S1–S3; Fig. 4).

Thus, we conclude that *CG6660* rather than *klg* is the causal gene that is associated with ethanol consumption. Similarly, we could confirm that an SNP in an intron of *grass* and downstream of *Men-b* implicated *Men-b* rather than *grass* as the causal gene (Tables S1–S3; Fig. 4).

Previous studies showed that many candidate genes affect the within-genotype environmental variance of sucrose consumption (Garlapow et al. 2015, 2016). Therefore, we estimated the effects of the candidate genes on the variance of ethanol and sucrose consumption relative to the appropriate control (Table S3) (Morgante et al. 2015). We identified only a few genes with a significant effect on the within-genotype variance compared to the control (*D19A*, *Mmp2*, *Trim9*, *sima*, *Men-b*, *Men-b*, and *slc45-1*).

Thus, we conclude that *CG6660* rather than *klg* is the causal gene that is associated with ethanol consumption. Similarly, we could confirm that an SNP in an intron of *grass* and downstream of *Men-b* implicated *Men-b* rather than *grass* as the causal gene (Tables S1–S3; Fig. 4).

Previous studies showed that many candidate genes affect the within-genotype environmental variance of sucrose consumption (Garlapow et al. 2015, 2016). Therefore, we estimated the effects of the candidate genes on the variance of ethanol and sucrose consumption relative to the appropriate control (Table S3) (Morgante et al. 2015). We identified only a few genes with a significant effect on the within-genotype variance compared to the control (*D19A*, *Mmp2*, *Trim9*, *sima*, *Men-b*, *Men-b*, and *slc45-1*).
Genetics of alcohol consumption in *Drosophila*

Figure 4: Genetic interaction networks for variation in ethanol consumption. (a) A network constructed from a subset of candidate genes identified by the extreme QTL mapping analysis at $P < 10^{-8}$. The network consists of 23 interconnected genes ($P < 0.001$). Blue font indicates genes with human orthologs. (b) A corresponding network of the human orthologs from (a), indicated in rectangular boxes with computationally recruited genes indicated in triangles.

Our results show that the genetic underpinnings for voluntary ethanol consumption are complex and different between males and females, in accordance with a sexually dimorphic architecture previously attributed to other alcohol-related phenotypes (Morozova et al. 2009, 2015).

**Genetic networks for voluntary ethanol consumption**

We asked to what extent the 291 genes with differentially segregating alleles among the extreme ethanol consumption groups (high vs. low, high vs. control, high vs. low + control) at a more lenient significance threshold of $P < 10^{-8}$ participate...
in known gene–gene interactions. We identified a network comprised of 23 interacting genes and a separate trio and a pair of connected genes (Fig. 4a). The probability that the larger network might have arisen by chance is $P < 0.001$. Gene ontology analysis of 23 interacting genes showed enrichment of developmental genes, including genes associated with development of the nervous system (Table S4). Thus, similar to other behavioral traits, naturally occurring phenotypic variation in voluntary ethanol consumption may in part be due to subtle variations in neuronal architecture.

Among the genes in the network, 86% have human orthologs. This enabled us to construct a human genetic interaction network based on the *Drosophila* genetic network associated with variation in ethanol consumption by allowing computational recruitment of two additional genes between connected orthologs (Fig. 4b). The resulting network comprised 32 genes, of which nine were orthologs corresponding to genes of the *Drosophila* network. Gene ontology enrichment analysis of this network highlighted signal transduction along with protein metabolic processes and development, including development of the nervous system ($P = 0.05$; Table S5).

**Discussion**

We used extreme QTL mapping to identify differentially segregating alleles in individuals with extreme ethanol consumption in a DGRP-derived AIP. The significant alleles tagged genes and a genetic network associated with natural variation in consumption of ethanol-supplemented sucrose. We administered ethanol in a sucrose solution to motivate consumption by reducing aversion to ethanol alone. We used mutational analyses to functionally implicate 86% of tested candidate genes in ethanol consumption in at least one sex, and to separate ethanol intake from sucrose intake per se. This level of validation represents an empirical false discovery rate similar to that obtained from validation studies on other phenotypes in GWA studies of the DGRP or extreme QTL mapping studies of DGRP-derived AIP (Garlapow et al. 2016; Morozova et al. 2015; Shorter et al. 2015; Swarup et al. 2013).

Quantitative RT-PCR of a sample of RNAi knockdown lines showed sexually dimorphic variation in the extent of reduction in target mRNA, similar to previous observations (Swarup et al. 2011; Fig. S2). In most cases, there was no correlation between the extent of RNAi knockdown and phenotypic effect, i.e. a small reduction in expression of a specific gene may elicit a large phenotypic effect and vice versa. Thus, when we observe a difference between the mutant and its co-isogenic control, we can infer an effect on the function of the target gene, no matter to what extent its transcript abundance is affected. It is of interest to note that quantitative RT-PCR analyses showed similar reduction in mRNA for *CG6660* and *klg* (Fig. S2), which were both tagged by the same SNP. However, only *CG6660* showed a phenotypic effect in both sexes (Fig. 3), providing convincing evidence that *CG6660* rather than *klg* is causally associated with variation in alcohol consumption.

Previous studies reported GWA analyses for variation in sucrose feeding in the DGRP and identification of candidate genes associated with high and low sucrose feeders in AIP-derived artificial selection lines (Garlapow et al. 2015, 2016). Not surprisingly, several of these genes are also contained in the network we identified (Fig. 3a), namely *rg*, *Egfr*, *jing*, *Fur1*, *simj*, *klg*, *fz2* and *fzd* (together ~29% of the genes in the network). However, an overall comparison between genes identified in these experiments and those identified in our study showed only ~27% overlap (Fig. S3), indicating that the predominant effect on variation in the phenotype in our study is due to ethanol. This is in agreement with findings by Sekhon et al. that about 30% of the common genes are contributing to both food and ethanol consumption (Sekhon et al. 2016).

Several genes identified in our extreme QTL mapping study have been associated previously with behavioral responses to ethanol and alcohol sensitivity, including *bun*, *cpr*, *CG13506*, *CG7386*, *CzR*, *dpr*, *Dhc98D*, *Egfr*, *fas*, *fzd*, *jing*, *klg*, *loci*, *msn* and *Mymo31DF* (Kong et al. 2010; Morozova et al. 2007, 2009, 2015). These genes are implicated in a wide range of biological processes, including development of the nervous system.

Fifty-eight genes (~20%) identified in GWA analyses across the DGRP for ethanol knockdown time (Morozova et al. 2015) correspond to genes identified in the present study, including *dpr*, *CG33158*, *jing*, *msn*, *Mmp2* and *rg*, indicating a continuum in the genetic architectures of feeding behavior, ethanol consumption and ethanol inebriation. It should be noted, however, that the AIP is derived from only 37 DGRP lines and contains 66% of the allelic variants present in the DGRP. Previous studies showed that despite differences in network compositions derived from GWA analyses in the DGRP and extreme QTL mapping in AIPs the genetic networks relate to the same cellular processes (Carbone et al. 2016; Morozova et al. 2015; Shorter et al. 2015; Swarup et al. 2013). We note that gene interactions and network architecture can vary based on developmental stage, tissue and environmental conditions.

We observed extensive sexual dimorphism in the genetic architecture for ethanol consumption. This is expected, as sexual dimorphism is a common feature of the genetic architecture of complex traits and has been documented for every sexual dimorphism is a common feature of the genetic architecture of complex traits and has been documented for every sex trait analyzed to date in the DGRP including but not limited to alcohol sensitivity (Morozova et al. 2015), sleep phenotypes (Harbison et al. 2013), olfactory behavior (Swarup et al. 2013), feeding behavior (Garlapow et al. 2015) and visual senescence (Carbone et al. 2016).

Gene ontology analyses of the *Drosophila melanogaster* genetic network associated with variation in ethanol consumption showed enrichment of developmental genes, including development of the nervous system (Table S4). Enrichment of neurodevelopmental genes has been a hallmark of genome-wide studies on variation in behavioral traits, such as olfactory behavior (Swarup et al. 2013) and susceptibility to environmental toxins (Zhou et al. 2016). When we translated the *Drosophila* genetic interaction network into an orthologous human network, 9 out of 30 input genes could be interconnected when we allowed two missing genes (i.e. genes not empirically implicated in association
with the trait). These genes anchored a network enriched in genes involved in signal transduction, protein metabolism and developmental processes, including nervous system development (Table S5).

In humans, alcohol consumption is a significant risk factor for cancer of the oral cavity and pharynx, esophagus, colorectum, liver, larynx and female breast (Bagnardi et al. 2015). Interestingly, our gene ontology enrichment analysis of human genes orthologous to the genes in the Drosophila gene network associated with ethanol consumption showed significant enrichment of the Kyoto Encyclopedia of Genes and Genomes (KEGG) prostate cancer pathway involving BAD, EGFR, MDM2, PDK1, PIK3CB and PIK3CA (Table S5). Furthermore, EGFR and ADAM10 may be responsible for coupling moderate alcohol consumption to breast cancer by increasing EGFR signaling (Mill et al. 2009). RAP1GAP has been associated with prostate (Bailey et al. 2009) and gastric cancers (Yang et al. 2016). EGFR and VVHAB have been previously mapped to a genetic network for alcohol-related phenotypes (Morozova et al. 2014).

In conclusion, we find that the genetic architecture that underlies variation in voluntary ethanol consumption is sexually dimorphic, partially overlaps and thus is distinct from, yet continuous with, genetic factors that control variation in feeding behavior and alcohol sensitivity, and contains evolutionarily conserved features that can be extrapolated to human genetic interaction networks. As is the case for other behavioral phenotypes, variations in neural connectivity and signaling appear to be associated with phenotypic variation in ethanol consumption.

References

Antonov, A.V. (2011) BioProfiling.de: analytical web portal for high-throughput cell biology. Nucleic Acids Res 39, W323–W327.
Antonov, A.V., Dietmann, S. & Mewes, H.W. (2008) KEGG spider: interpretation of genomics data in the context of the global gene metabolic network. Genome Biol 9, R179.
Bagnardi, V., Rota, M., Tramacere, I., Islami, F., Fedirko, V., Bailey, C.L., Kelly, P. & Casey, P.J. (2009) Activation of Rap1 promotes development (Table S5).
Bier, E. (2005) Drosophila, the golden bug, emerges as a tool for human genetics. Nat Rev Genet 6, 9–23.
Bierut, L.J., Agrawal, A., Bucholz, K.K. et al. (2010) A genome-wide association study of alcohol dependence. Proc Natl Acad Sci USA 107, 5082–5087.
Bierut, L.J., Goate, A.M., Breslau, N. et al. (2012) ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Mol Psychiatry 17, 445–450.
Carbone, M.A., Yamamoto, A., Huang, W., Lyman, R.A., Meadors, T.B., Yamamoto, R., Anholt, R.R.H. & Mackay, T.F.C. (2016) Genetic architecture of natural variation in visual senescence in Drosophila. Proc Natl Acad Sci USA 113, E6620–E6629.
Clarke, T.K., Smith, A.H., Gelernter, J., Kranzler, H.R., Farber, L.A., Hill, L.S., Fernandez-Pujals, A.M., MacIntyre, D.J., Smith, B.H., Hocking, L.J., Padmanabhan, S., Hayward, C., Thompson, P.A., Porteous, D.J., Deary, I.J. & McIntosh, A.M. (2016) Polygenic risk for alcohol dependence associates with alcohol consumption, cognitive function and social deprivation in a population-based cohort. Addict Biol 21, 469–480.
Dennis, G. Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C. & Lempicki, R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4, R3.
DePristo, M.A., Banks, E., Poplin, R., Gabriella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., de Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kerns, T.S., Aivachencenco, A.Y., Cibuls, K., Gabriel, S.B., Altshuler, D. & Daly, M.J. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491–498.
Dietz, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Felliner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keelman, K. & Dickson, B.J. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature 448, 151–156.
Edenberg, H.J. & Foroud, T. (2013) Genetics and alcoholism. Nat Rev Gastroenterol Hepatol 10, 487–494.
Edenberg, H.J., Koller, D.L., Xue, X. et al. (2010) Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res 34, 840–852.
Ehrenreich, I.M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J.A., Gresham, D., Caudy, A.A. & Kruglyak, L. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464, 1039–1042.
Frank, J., Cichon, S., Treutlein, J. et al. (2012) Genome-wide significant association between alcohol dependence and a variant in the ADH gene cluster. Addict Biol 17, 171–180.
Garlapow, M.E., Huang, W., Yarboro, M.T., Peterson, K.R. & Mackay, T.F.C. (2015) Quantitative genetics of food intake in Drosophila melanogaster. PLoS One 10, e0138129.
Garlapow, M.E., Everett, L.J., Zhou, S., Gearhart, A.W., Fay, K.A., Huang, W., Morozova, T.V., Wu, G.H., Turlapati, L., St. Armour, G., Hussain, Y.N., Mcdamms, S.E., Fochler, S. & Mackay, T.F.C. (2016) Genetic and genomic response to selection for food consumption in Drosophila melanogaster. Behav Genet 47, 227–243.
Gelernter, J., Kranzler, H.R., Sherva, R., Almasy, L., Koesterer, R., Smith, A.H., Anton, R., Preuss, U.U., Ridinger, M., Rujescu, D., Wodarz, N., Zill, P., Zhao, H. & Farber, L.A. (2014) Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. Mol Psychiatry 19, 41–49.
Grueber, W.B., Ye, B., Yang, C.H., Younger, S., Borden, K., Jan, L.Y. & Jan, Y.N. (2007) Projections of Drosophila multidendritic neurons in the central nervous system: links with peripheral dendrite morphology. Development 134, 65–64.
Harbison, S.T., McCoy, J.J. & Mackay, T.F.C. (2013) Genome-wide association study of sleep in Drosophila melanogaster. BMC Genomics 14, 281.
Heath, A.C., Whitfield, J.B., Martin, N.G., Pergadia, M.L., Goate, A.M., Lind, P.A., McEvoy, B.P., Schrage, A.J., Grant, J.D., Chou, Y.L., Zhu, R., Henders, A.K., Medland, S.E., Gordon, S.D., Nelson, E.C., Agrawal, A., Nyholt, D.R., Bucholz, K.K., Madden, P.A. & Montgomery, G.W. (2011) A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. Biol Psychiatry 70, 513–518.
Genetics of alcohol consumption in Drosophila with epithelial-mesenchymal transition (EMT) and poor prognosis in gastric cancer. Oncotarget 8, 8057–8068.
Yonekura, S., Xu, L., Ting, C.Y. & Lee, C.H. (2007) Adhesive but not signaling activity of Drosophila N-cadherin is essential for target selection of photoreceptor afferents. Dev Biol 304, 759–770.
Zhou, S., Morozova, T., Hussain, Y.N., Luoma, S.E., McCoy, L., Yamamoto, A., Mackay, T.C. & Anholt, R.R.H. (2016) The genetic basis for variation in sensitivity to lead toxicity in Drosophila melanogaster. Environ Health Perspect 124, 1062–1070.

Acknowledgments
This work has been supported by NIH grants R01 AA016560 and R01 GM059469 to T.F.C.M. and R.R.H.A.

Supporting Information
Additional supporting information may be found in the online version of this article at the publisher’s web-site:

Figure S1. Power to detect allele frequency difference in extreme QTL mapping. To show the statistical power to detect allele frequency difference, we used simulation to derive the power. We assumed the same test statistic as described in the main text, a P-value threshold according to Bonferroni correction (2.14 x 10^{-8}), 100 flies were sequenced in each pool, at a sequence coverage of 209x (median coverage in all tested sites in our dataset). Power is calculated as the number of simulations from a total of 100,000 in which the assumed allele frequency difference produces a significant result.

Figure S2. Relative fold changes in mRNA levels of target genes for Ubiquitin-GAL4/UAS-RNAi lines compared to the co-isogenic control (60010). The dashed line indicates the expression level in the control line. Red bars designate females and blue bars designate males. Error bars are ± SEM. *P < 0.05; **P < 0.001; ***P < 0.0001; Student's t-test. PCR primer sequences are given in Table S6.

Figure S3. Overlap between candidate genes identified in extreme QTL mapping for ethanol consumption (this study) and sucrose consumption (Garlapow et al. 2015, 2016).

Table S1. Extreme QTL mapping analysis in an outbred advanced intercross Drosophila population. The frequency of the major allele is given for each replicate and the average of all replicates.

Table S2. Candidate genes associated with ethanol consumption from extreme QTL mapping analysis chosen for functional analyses. H, high ethanol consumers; L, low ethanol consumers; C, control group; L+C, low and control groups combined. Only the most significant SNP for each gene is indicated.

Table S3. Functional analyses of candidate genes associated with ethanol consumption.

Table S4. Gene ontology enrichment analysis for 23 interconnected Drosophila candidate genes (enrichment scores ≥2.5).

Table S5. Gene Ontology enrichment analyses for human orthologs of Drosophila candidate genes in the genetic network (enrichment scores ≥2.5).

Table S6. Quantitative RT-PCR primer sequences.

Genes, Brain and Behavior (2017) 16: 675–685