Analysis of natural variation reveals neurogenetic networks for *Drosophila* olfactory behavior

Shilpa Swarup[a,b,1,2], Wen Huang[a,b,1], Trudy F. C. Mackay[a,b,3], and Robert R. H. Anholt[a,b,c,3]

Departments of [a]Genetics and [c]Biology and [b]W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695-7614

Contributed by Trudy F. C. Mackay, November 27, 2012 (sent for review September 28, 2012)

Understanding the relationship between genetic variation and phenotypic variation for quantitative traits is necessary for predicting responses to natural and artificial selection and disease risk in human populations, but is challenging because of large sample sizes required to detect and validate loci with small effects. Here, we used the inbred, sequenced, wild-derived lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP) to perform three complementary genome-wide association (GWA) studies for natural variation in olfactory behavior. The first GWA focused on single nucleotide polymorphisms (SNPs) associated with mean differences in olfactory behavior in the DGRP, the second was an extreme quantitative trait locus GWA on an outbred advanced intercross population derived from extreme DGRP lines, and the third was for SNPs affecting the variance among DGRP lines. No individual SNP in any analysis was associated with variation in olfactory behavior by using a strict threshold accounting for multiple tests, and no SNP overlapped among the analyses. However, combining the top SNPs from all three analyses revealed a statistically enriched network of genes involved in cellular signaling and neural development. We used mutational and gene expression analyses to validate both candidate genes and network connectivity at a high rate. The lack of replication between the GWA analyses, small marginal SNP effects, and convergence on common cellular networks were likely attributable to epistasis. These results suggest that fully understanding the genotype–phenotype relationship requires a paradigm shift from a focus on single SNPs to pathway associations.

genetic architecture | chemosensation | behavioral genetics

Understanding the rules by which variation in primary DNA sequence impacts variation for quantitative traits in natural populations is critical for predicting responses to natural and artificial selection and disease risk in human populations. The emerging picture from large genome-wide association (GWA) studies in human populations is that many common variants with individually small marginal (additive) effects affect diseases and quantitative traits, of which only a small fraction can be replicated across populations (1, 2). Although it is possible that effects of common single nucleotide polymorphisms (SNPs) underestimate the true effects because causal variants are not common or not SNPs, and lack of replication is due to differences in allele frequency and pattern of linkage disequilibrium (LD) (1, 2), it is also possible that small additive effects and lack of replication are due to underlying epistatic interactions (3, 4).

Extremely large samples are required to determine individual significance of rare alleles and epistatic interactions. One approach to gaining biological insight from GWA studies in the absence of statistical significance of individual SNPs is functional evaluation of genes harboring the top SNPs regardless of individual significance. A second approach is to consider the genes with the top signals collectively and examine whether they are enriched for known pathways (5–8). Both approaches test the hypothesis that the *P* values did not reach formal significance because of imperfect LD with true casual variants and/or unmodeled epistatic interactions. Here, we evaluated the feasibility of these strategies by using the *Drosophila melanogaster* Genetic Reference Panel (DGRP), a collection of inbred wild-derived lines with fully sequenced genomes (9). The DGRP enables GWA studies where all variants are known, environmental conditions can be controlled, and phenotyping many genetically identical individuals of each line gives an accurate estimate of the genotypic effect. Large publicly available collections of mutations and RNAi constructs (10) facilitate rapid and economical functional evaluation of candidate genes, allowing empirical assessment of the false discovery rate (FDR); outbred populations derived from DGRP lines can be constructed for multilocus evaluation of allelic effects (11).

We focused on olfactory behavior—an important quantitative trait because behavioral responses to chemical signals are essential for survival and reproduction. The *D. melanogaster* olfactory system is well characterized (12), with four multigene chemosensory families comprising *Odorant receptor* (*Or*) (13–16), *Ionotropic receptor* (*Ir*) (17), *Gustatory receptor* (*Gr*) (18, 19), and *Odorant binding protein* (*Obp*) (20, 21) genes. Mutational analyses implicate many other loci required for mediating olfactory behavior in addition to peripheral chemoreceptors (22, 23), and these mutations form genetic interaction networks (23, 24). There is substantial natural variation for olfactory behavior (25–27). Candidate gene-focused association analyses indicate that this variation is, in part, attributable to polymorphisms in *Obp* (26–28) and *Or* (29) genes. However, addressing the question about contributions of other genes involved in development and central processing of signals to natural variation in odor-guided sensorimotor integration has become possible with the generation of the DGRP.

We performed three GWA analyses to identify alleles that affect natural variation in olfactory behavior. First, we performed a GWA on DGRP lines to identify top SNPs associated with mean differences in olfactory behavior. Second, we intercrossed two divergent DGRP lines and constructed an outbred advanced intercross line (AIL) population, on which we performed an extreme quantitative trait locus (QTL) GWA (30). Finally, we performed a GWA for SNPs affecting the variance among lines (variance GWA, vGWA) as a one-dimensional screen for interacting loci (31, 32), because there is evidence for epistasis between mutations affecting olfactory behavior (23, 24) and mutational effects are suppressed in DGRP line genetic backgrounds (33). We find that different elements of the genetic architecture that underlies natural variation in olfactory behavior are revealed in the three GWA analyses, but they converge on similar cellular processes associated with neural signaling and neural connectivity, and are functionally validated at a high rate.

Results

GWA Analysis for Olfactory Behavior. We measured olfactory behavior to benzaldehyde for 164 DGRP lines with a modification

Author contributions: T.F.C.M. and R.R.H.A. designed research; S.S. performed research; S.S. and W.H. analyzed data; and S.S., W.H., T.F.C.M., and R.R.H.A. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]S.S. and W.H. contributed equally to this work.

[2]Present address: Genome Biology Team, Monsanto Company, Chesterfield, MO 63017.

[3]To whom correspondence may be addressed. E-mail: trudy_mackay@ncsu.edu or anholt@ncsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1220168110/-/DCSupplemental.

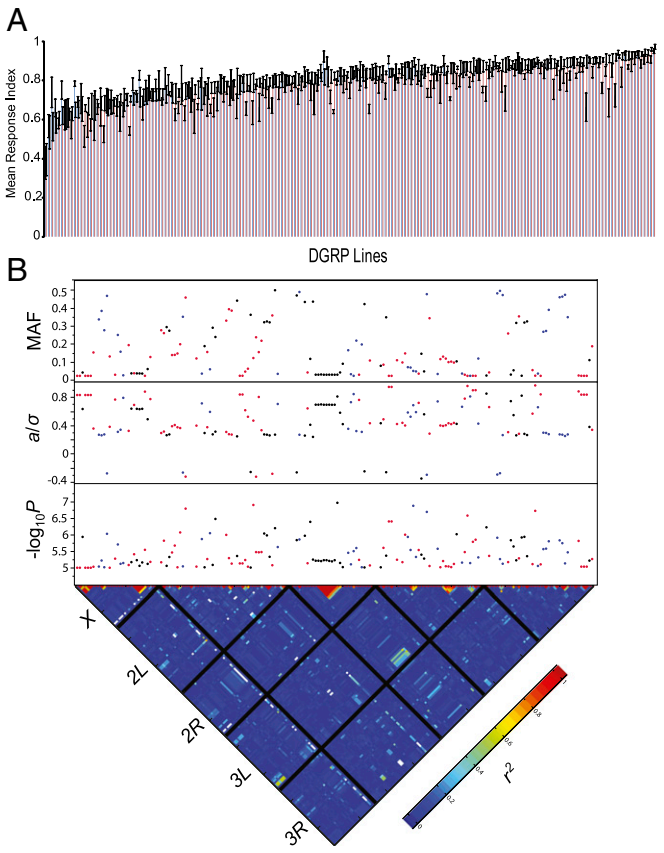

Fig. 1. Association between SNPs in the DGRP and variation in olfactory behavior. (*A*) Line means for olfactory behavioral response to 0.3% (vol/vol) benzaldehyde for DGRP males (blue bars) and females (pink bars). Error bars are SEMs. (*B*) Genome-wide association analyses. All SNPs from single-marker analyses with $P < 10^{-5}$ are shown. (*Upper*) Associations based on females are indicated by red dots, males by blue dots, and sexes pooled by black dots. (*Lower*) The triangle depicts LD between SNPs as measured by r^2, with the five major chromosome arms demarcated by black lines. The heat map indicates the magnitude of LD with red corresponding to complete LD and blue to absence of LD. *Upper* shows the significance threshold ($-\log_{10}P$), the effect size in phenotypic SD units (a/σ_P), and the minor allele frequency (MAF).

(34) of the "dipstick" assay (22). We found substantial phenotypic and genetic variation in olfactory behavior, with a broad sense heritability of $H^2 = 0.45$ (Fig. 1*A* and Table S1). As observed previously (22, 25–27), there was sexual dimorphism in the behavioral response to benzaldehyde and genetic variation in the magnitude of sex dimorphism, with a cross-sex genetic correlation of $r_{MF} = 0.91$) (Fig. 1*A* and Table S1). Thus, we expect that many alleles affecting variation in olfactory behavior will be common to males and females, but that some alleles will have sex-specific effects.

We performed single-marker GWA analyses pooled across sexes and separately for males and females for 2,587,691 SNPs. None of the SNPs were significantly associated with variation in olfactory behavior at a Bonferroni-corrected *P* value of 1.9×10^{-8} in any analysis. To examine whether the top SNPs were enriched for true positive associations and cellular networks, we lowered the discovery significance threshold to $P < 10^{-5}$. At this *P* value, there were 184 unique SNPs (79 SNPs for females, 51 for males, 66 for sexes pooled, and 35 for the SNP by sex interaction) associated with variation for olfactory behavior (Fig. 1*B* and Dataset S1). Lower frequency alleles had larger effects than common alleles (Fig. 1*B*), consistent with GWA studies on other complex traits in this population (9, 35, 36).

The SNPs that are associated with variation in olfactory response to benzaldehyde were located in or near 176 genes, of which ~40% encode predicted transcripts of unknown function (Dataset S1). Many genes are involved in signal transduction (*Ac76E*, *Fak56D*, *inaC*, *IP3K1*, *Pde9*, *Pde1c, Pkc53E*) and development and function of the nervous system (*5-HT7*, *Atet*, *btl*, *CadN*, *CcapR*, *elk*, *exex*, *fw*, *Fas2*, *mam*, *pigs*, *PsGEF*, *Sdc*, *tkv*, *tutl*, *tws*). Several chemoreceptor genes not implicated previously in response to benzaldehyde harbor SNPs associated with phenotypic variation in response to this odorant in the DGRP. These genes include *ppk11*, which encodes a sodium channel implicated in gustatory response to salt (37), gustatory receptor *Gr92a,* and odorant receptor *Or49a*. The molecular response profiles of *Gr92a* and *Or49a* have not been reported. Previous studies showed that mutations in *Spn* affect olfactory behavior to benzaldehyde (23), and mutations in *Spn* and *Btk29A* affect intermale aggressive behavior (38), which has an olfactory component.

Functional Validation and Network Analysis. We selected 18 candidate genes identified by GWA for functional validation, based on SNP location (within 1 kb upstream or downstream from the coding region, or nearest gene), strength of significance (*P* value), and availability of *P{MiET1}* insertion mutants (39). Nine mutants had significant effects on olfactory behavior in both sexes, one mutant had a female-specific effect, and four mutants had male-specific effects (Fig. 2). Thus, 78% of candidate genes that harbor SNPs nominally associated with variation in olfactory behavior in the GWA were functionally validated through mutational analysis, supporting the contention that the top SNPs were enriched for true positive associations.

Next, we performed a network enrichment analysis (40) to assess to what extent gene products associated with variation in olfactory behavior might interact in cellular pathways. Allowing for two missing genes, we identified a significantly enriched network ($P < 0.01$) comprised of 23 genes associated with cGMP metabolism, inositol triphosphate signaling, and neural development (Fig. 3*A*).

Five of the 18 *P{MiET1}* insertion mutants tested for effects on olfactory behavior (*Fak56D*, *Btk29A*, *Pkc53E*, *Pde1c*, *Ac76E*) belonged to this network, and all showed altered responses to benzaldehyde (Fig. 2). To further validate connectivity of the predicted network, we chose one of the most highly connected

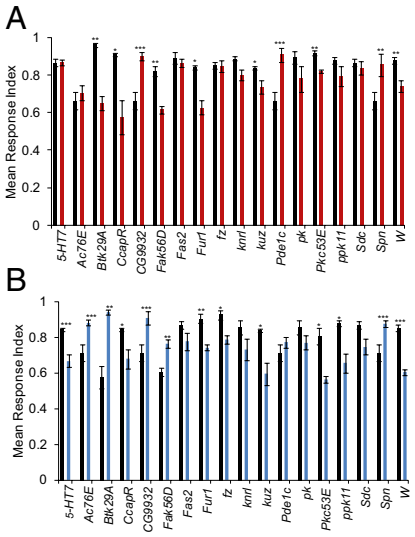

Fig. 2. Validation of candidate genes using mutants. Candidate genes from the GWA analyses were confirmed by using *P{MiET1}* mutants. Eighteen mutants were tested for females (*A*, red bars) and males (*B*, blue bars) along with their corresponding controls (black). Error bars indicate the SEMs. *$0.01 \le P \le 0.05$; **$0.001 \le P \le 0.01$; ***$0.0001 \le P \le 0.001$.

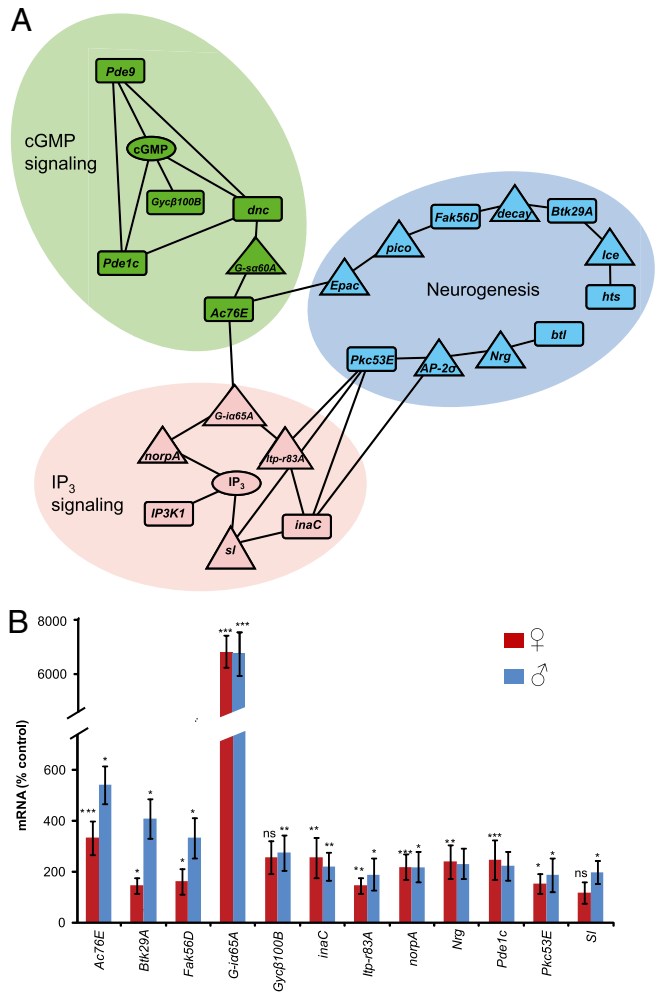

Fig. 3. Functional interactions among candidate genes. (*A*) An enriched ($P < 0.01$) cellular network (31) among candidate genes with at least one SNP ($P < 10^{-5}$) in the GWA analysis. Candidate genes are indicated by rectangles, missing genes (i.e., genes without significant associations) by triangles, and metabolites by circles. Components of the network associated with distinct interconnected cellular processes are highlighted by the colored backgrounds. (*B*) Validation of the connectivity of the predicted network by qRT-PCR of transcripts in the *Pkc53e P{MiET1}*-insertion mutant. Error bars indicate the SEMs. *$0.01 \le P \le 0.05$; **$0.001 \le P \le 0.01$; ***$0.0001 \le P \le 0.001$.

genes, *Pkc53E,* as the focal gene and measured transcript levels of 12 candidate genes connected directly or indirectly to *Pkc53E*. The *P{MiET1}* insertion in *Pkc53E* resulted in up-regulation of the corresponding transcript (Fig. 3*B*). Quantitative RT-PCR (qRT-PCR) analysis showed that transcript levels of all genes in the network, except *Pde1c* and *Nrg* for males and *Gycβ100B* and *sl* for females, were significantly up-regulated in the *Pkc53E P{MiET1}* mutant compared with the corresponding control (Fig. 3*B*), indicating coordinated regulation of their expression. Thus, our results show that neuronal signal transduction networks are associated with variation in olfactory behavior.

Extreme QTL Mapping. Because the GWA analysis was underpowered for less common variants in the DGRP, we adopted a complementary approach and constructed reciprocal AIL populations from two DGRP lines with divergent olfactory responses to benzaldehyde and performed extreme QTL mapping (24) in the F28 generation. We sequenced DNA pools of the 10% high and low tails of the distribution of response scores and compared frequencies of alleles that differed in the two parental lines. We identified 377 and 431 SNPs with a difference in allele frequency ≥0.2 at $P < 0.001$ between the top and bottom pools for females and males, respectively (Fig. S1 and Datasets S2 and S3). These SNPs were distributed within or near 417 (females) and 472 (males) genes across the genome, indicating a highly polygenic genetic architecture (Datasets S2 and S3). Alleles from the high-scoring line did not always increase avoidance to benzaldehyde in the AIL population, possibly due to context-dependent allelic effects or stabilizing selection for olfactory behavior preventing accumulation of alleles with effects in the same direction in either parental line.

We chose *P{MiET1}* insertion mutants corresponding to candidate genes identified by extreme QTL mapping to test for effects on olfactory behavior based on several criteria. SNPs in the candidate genes had low *P* values and high allele frequency differences between the extreme pools, were located in or near the gene, and the genes were expressed in the fly head. We tested 7 mutants for males and 11 for females (Fig. S2). For females, six mutants showed reduced responses to benzaldehyde, including *A2bp1*, a DNA binding protein involved in positive regulation of transcription (41); *Beat-Ia*, which is involved in axon guidance (42); and *smooth*, which is implicated in axon guidance and mating behavior (43). Mutations in three genes, *CG15144*, *Cpr92F*, and *Pvf3*, showed increased response indices. In males, six mutants, including *chinmo,* which is involved with mushroom body development and dendrite morphogenesis (44), and *sima*, implicated in intracellular signaling (45), showed significantly higher responses to benzaldehyde compared with the control. Thus, although the power to detect allele frequency differences was low given the sequencing depth, 83% of candidate genes could be validated through mutational analysis, similar to the 78% success rate found for genes identified by GWA in the DGRP.

Identification of Variance-Associated SNPs. We next performed GWA with line variance (variance GWA or vGWA). This analysis complements traditional GWA for SNPs associated with differences in trait mean and is a one-dimensional screen for interacting loci (31, 32). For each of the ~2.5 million SNPs, we performed a Levene's test (46) to compare variances among lines carrying the major allele and those carrying the minor allele. We found 21 and 14 SNPs associated with variance ($P < 10^{-5}$) in females and males, respectively, located within or near 22 (female) and 16 (male) genes (Dataset S4), with functions that again related to neural development and neurotransmitter signaling. Among these genes were five transcriptional regulators (*odd*, *crol*, *kni*, *CG1677*, and *sbb*), four transmembrane transporters (*CG31689*, *List*, *CG15096*, and *CG13646*), the nicotinic acetylcholine receptor β subunit and the glycine receptor, an adenylate cyclase (*CG43373*), and *Chemosensory protein A75a*.

GWA, vGWA, and AIL Analyses Uncover Different Elements of a Common Cellular Network. The three analyses identified different SNPs. Differences between GWA and vGWA results are expected, because GWA is designed to detect mean shifts, whereas vGWA contrasts variances of the two genotypes at each SNP, as expected with epistatic interactions in which the effect of an undetected interacting locus has a much greater effect in one genetic background than the other at the tested locus. Discordance between SNPs identified by GWA of differentially segregating alleles in the AIL populations can be attributed to context-dependent effects (11) and variants that are not common in the DGRP but at intermediate frequencies in the AIL population. To exclude possibilities, other than epistasis, that could result in discrepancies between these complementary approaches, we verified that there was no enrichment of significant SNPs in the AIL that could not be reliably tested by GWA if they were present at low frequency in the DGRP (i.e., fewer than four lines carrying the minor allele) (Fig. S2). Relaxing the significance threshold in the GWA analysis did not recover more significant SNPs in extreme QTL mapping than selecting SNPs randomly (Fig. S3). SNPs in close physical proximity to GWA signals in the DGRP were more strongly associated

with the trait than randomly selected SNPs in the DGRP (Fig. S4*A*), as expected from local LD, but this was not observed for DGRP SNPs in close proximity to AIL signals (Fig. S4*B*). Conversely, SNPs near the top AIL SNPs were more strongly associated with the trait than randomly selected SNPs in the AIL population, but not around DGRP signals (Fig. S4 *C* and *D*). Thus, the top SNPs detected by GWA analysis for the trait mean in the DGRP and by extreme QTL mapping in the AIL population are unlikely to tag the same genomic regions. Similar observations have been made in a separate synthetic outbred population derived from the DGRP lines for three other traits (11).

If GWA results depend on genetic context, an analysis combining the results for different genetic backgrounds should converge on a common underlying genetic network. Therefore, we pooled all candidate genes harboring SNPs associated with phenotypic variation in olfactory behavior from GWA, vGWA, and AIL analyses and performed a network enrichment analysis (40). We uncovered a significant ($P < 0.005$) cellular network allowing for one missing gene (Fig. 4) that expanded the network identified by using GWA data alone (Fig. 3*A*). This network comprised genes associated with axon guidance, IP3 signaling, cGMP metabolism, cell adhesion, and neural development (Fig. 4), implying that polymorphisms may contribute to natural variation in olfactory perception through subtle variations in neuronal signaling and connectivity of the nervous system.

Although there were no SNPs in common between the GWA analysis in the DGRP and the AIL population, the SNPs tagged 22 common genes: *5-HT7*, *aret*, *Atet*, *beat-VI*, *btl*, *Btk29A*, *CadN*, *CG7737*, *CG31262*, *CG42747*, *CR43990*, *Eip93F*, *elk*, *Fak56D*, *Fur1*, *fz*, *Gr92a*, *mir-276a*, *Pkc53E*, *Sox21b*, *tutl*, *tws*). We tested mutations in seven of these genes (*5-HT7*, *Btk29A*, *elk*, *Fak56D*, *Fur1*, *fz*, *Pkc53E*) for effects on olfactory behavior, and all seven validated (Fig. 2 and Fig. S1). Further, four of the genes in common (*Btk29A*, *btl*, *Fak56D*, *Pkc53E*) were in the enriched network derived by combining all three analyses.

Discussion

Before the availability of complete genome sequences for the DGRP lines, association studies for olfactory behavior focused on a few *Or* and *Obp* genes (26–29). Here, we identified SNPs associated with variation in olfactory behavioral response to benzaldehyde across the entire genome using three different GWA analyses. None of the SNPs in any analysis were significant at a Bonferroni multiple testing threshold, nor did any SNP overlap among the analyses (although 22 genes were in common). Prior GWA analyses for starvation resistance (9), chill coma recovery time (9), startle response (9), and acute (35) and chronic (36) responses to oxidative stress similarly detected many loci associated with each trait at a discovery P value $<10^{-5}$. Although none of the trait-associated SNPs in these studies exceeded Bonferroni significance thresholds, implicated genes had high validation rates and recapitulated known cellular networks not previously associated with the traits (35, 36). Attempts to replicate multi-SNP associations in an outbred population derived from DGRP lines

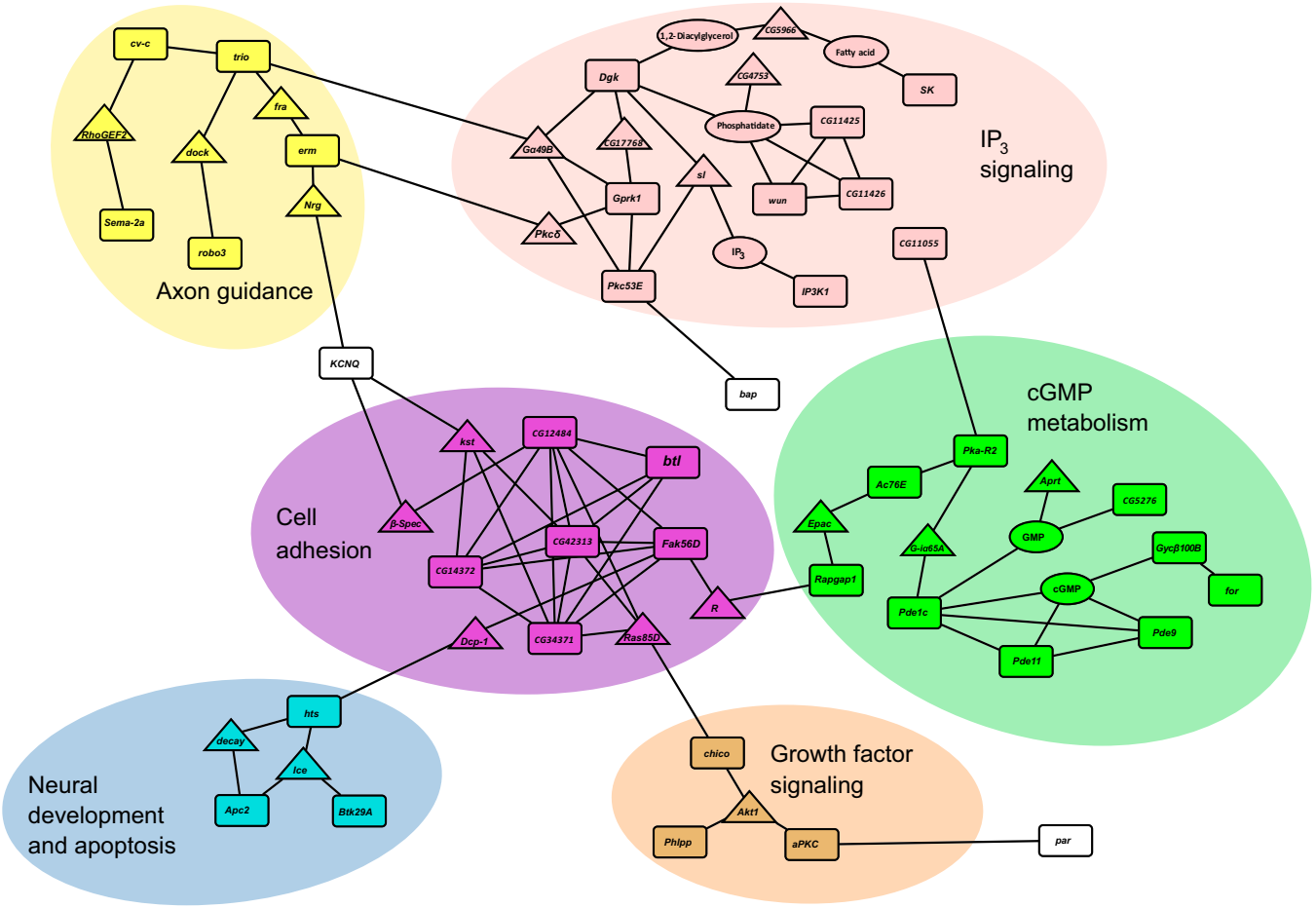

Fig. 4. Cellular network for olfactory behavior identified by combined GWA, vGWA, and AIL analyses. An enriched ($P < 0.005$) cellular network (31) allowing for one missing gene derived from genes containing at least one SNP within 1 kb in any of the GWA, vGWA, and AIL analyses ($P < 10^{-5}$). Significant genes are indicated by rectangles, missing genes by triangles, and metabolites by circles. Components of the network associated with distinct interconnected cellular processes are highlighted by the colored backgrounds.

failed, but the majority of the SNPs detected at Bonferroni-corrected significance levels in the outbred population participated in at least one epistatic interaction in the DGRP (11).

The null hypothesis for the lack of significant associations and replication in the three GWA analyses for olfactory behavior is that all are false positive associations due to the lenient discovery significance thresholds. The alternative hypothesis is that these top signals are enriched for true positive associations because of imperfect LD with causal variants (low frequency SNPs or non-SNP variants) and/or unmodeled epistatic interactions. We showed that 29 of 36 tested mutations (80.6%) implicated by the DGRP and extreme QTL mapping GWA analyses affected olfactory behavior, a much greater proportion than the 6% (22, 23) expected from random *P*-element insertional mutations (Fisher's exact test; $P = 4.79 \times 10^{-11}$). Furthermore, the genes implicated by SNP associations in the DGRP were statistically enriched for a network of genes centered on signal transduction pathways and neural connectivity. *Pkc53E* was a highly connected gene. A mutation in *Pkc53E* affected olfactory behavior, and comparison of gene expression in the mutant and its coisogenic control validated the network connectivity at the level of gene expression. This network was expanded when genes tagged by the top SNPs in all three GWA analyses were considered jointly.

Our results suggest that polymorphisms that contribute to natural variation in olfactory perception are not restricted to peripheral chemoreceptors, but may also cause subtle variations in genes affecting neural connectivity and signaling in the olfactory projection. These observations are in line with previous studies, which used optical recording techniques to document odorant-specific activation patterns in the mushroom bodies (47–49) and demonstrated anatomically distinct projections from individual glomeruli to the lateral horn of the protocerebrum (50). Furthermore, signaling via both $G\alpha_q$ and $G\alpha_s$ in the mushroom bodies have been implicated in discrimination of unconditioned odor identity and experience-dependent odor discrimination, respectively (49), consistent with the observed prominence of both inositol-triphosphate and cyclic nucleotide signaling pathways in our GWA analyses. Thus, variation in neural connectivity and synaptic signaling in the central nervous system is a major determinant of individual variation in odor perception.

Biological networks are characterized by many positive and negative feedback loops leading to nonlinear interactions among gene products (51). Arguably, epistatic interactions among genetic variants perturbing these networks are more biologically plausible than additive interactions, and additivity is likely an emergent property of variation in underlying epistatic networks. Epistasis can lead to substantial underestimates of main effects of SNPs participating in an epistatic interaction, if the interacting partner and the interaction term are not included in the statistical model used to infer association. Epistasis is prevalent among *P*-element mutations affecting olfactory behavior (23, 24), and the DGRP lines harbor variants that suppress the effects of these mutations (33). Epistatic interactions will lead to the lack of replication of single marker SNP associations with a quantitative trait in populations with different allele frequencies, but the underlying biological interaction networks can be inferred from combining GWA results from such populations (11, 52). These considerations lead to a paradigm shift in the way we view the relationship between genetic and phenotypic variation. There may be much less phenotypic variation in natural populations than expected given the large amount of segregating genetic variation, due to the evolution of canalization via suppressing epistatic interactions between common alleles (33). This hypothesis is testable in model organisms and is one possible explanation for the problem of missing heritability in human GWA studies (1).

Methods

***Drosophila* Stocks.** The DGRP was generated by 20 generations of full sib mating of progeny of inseminated females from the Raleigh, NC, population (9). *P{MiET1}* mutants and coisogenic controls (39) were obtained from the Bloomington *Drosophila* stock center. Flies were reared on cornmeal-molasses-agar-yeast medium at 25 °C, 70% humidity, and a 12-h light/dark cycle.

Behavioral Assay. Olfactory behavior of single-sex groups of 50 flies per replicate and three replicates per sex to 0.3% (vol/vol) benzaldehyde (Sigma-Aldrich) was measured for each line between 1400 and 1600 hours, using a modification (34) of the "dipstick assay" (22). A response index (RI) of 1 indicates a maximal aversive response to the odorant, and RI = 0 indicates that all flies remain near the odor source. Replicates were run on different days to randomize environmental variation. Data were analyzed with the ANOVA model $Y = \mu + L + S + L \times S + \varepsilon$ to partition phenotypic variance among lines (*L*, random), sex (*S*, fixed), line by sex interaction ($L \times S$, random), and within-line variance (ε). Broad-sense heritabilities and the cross-sex genetic correlation for olfactory behavior were estimated from the variance components as $H^2 = (\sigma_L^2 + \sigma_{SL}^2)/(\sigma_L^2 + \sigma_{SL}^2 + \sigma_E^2)$ and $r_{MF} = \sigma_L^2/(\sigma_L^2 + \sigma_{SL}^2)$, respectively.

Genome-Wide Association Analyses for Means and Variances. The 2,587,691 SNPs used for association analyses had minor alleles present in at least four DGRP lines; had coverage between 2 and 30; were biallelic and not called segregating within a line; and were genotyped in at least 60 DGRP lines. Single marker tests for association with variation in line means were performed pooled across sexes using the ANOVA model $Y = \mu + M + S + M \times S + L(M) + \varepsilon$, and separately for males and females using the ANOVA model $Y = \mu + M + \varepsilon$, where μ is the overall mean, *M* is the SNP effect, *S* is sex, *L* is line and ε is the error variance. These analyses were performed using SAS software. To identify SNPs that were associated with different variances among lines for olfactory behavior, tests for homogeneity of variance between the two genotypes were performed by using Levene's test (46), implemented in the *car* package in R.

Assessment of Gene Expression Levels. Transcript levels were assessed by qRT-PCR using a SYBR green detection method (Applied Biosystems) with glyceraldehyde-3-phosphate dehydrogenase as the internal standard. Independent triplicates of total RNA were extracted from whole males and females ($n = 15$ per replicate) using TRIzol reagent (Invitrogen). cDNA was generated from 80 to 100 ng of total RNA by reverse transcription, and each extract was analyzed in duplicate. Transcript specific primers were designed to amplify 100- to 150-bp fragments. Negative controls without reverse transcriptase were run to exclude genomic contamination. Statistically significant differences in gene expression levels between *P{MiET1}* mutants and their controls were determined by two-tailed *t* tests.

Bioinformatics Analysis. Annotation of SNPs was based on Flybase release 5.46 (53). SNPs were considered in a gene if they were located in or within 1 kb upstream and downstream of a gene model. SNPs not located within 1 kb of an annotated gene were assigned to the nearest upstream and/or downstream genes within a window of 50 kb. The R-spider program in the BioProfiling.de web portal was used to identify ensembles of interacting gene products (40). This analysis tool incorporates data for ~2,000 genes and combines signaling and metabolic pathways from Reactome and KEGG databases to determine whether interactions between the input genes are greater than expected by chance. The network is built by connecting genes with known interactions in the two databases, allowing zero, one, or two missing nodes. The significance of the network is tested by Monte Carlo simulations, in which the same number of randomly selected genes is used to form the null distribution of the size of the network.

Extreme QTL Mapping. We generated reciprocal AIL mapping populations to perform extreme QTL mapping (30). Reciprocal AIL were initiated by crosses between DGRP lines RAL_357 (a low responder) and RAL_820 (a high responder) and were subsequently maintained by random mating in large populations each generation. Between generations 28 and 30, the top ($n = 200$) and bottom ($n = 200$) 10% responders to 0.3% (vol/vol) benzaldehyde were collected for each sex separately from each AIL population. These individuals were retested the next day to verify response reproducibility and pooled. DNA was extracted from two biological replicates of high and low pools, purified by using Genomic-tip 100/G columns (Qiagen), and sequenced on the Illumina Genome Analyzer GA_{iix} using 75 bp paired end reads. Each sample was sequenced in two lanes. Sequencing reads were mapped to the *D. melanogaster* reference genome by using BWA (0.6.1-r104; ref. 5), allowing a maximum of four mismatches and with the option "-q 13" to dynamically trim low quality ends of sequence. Mapped reads were locally realigned and marked for duplicates using GATK version 1.5–20 (54) and Picard tools (version 1.60, http://picard.sourceforge.net). Quality scores of bases were recalibrated using GATK. We counted $Q \geq 13$ bases

supporting the two parental alleles at segregating sites and compared allele frequencies in the top and bottom pools using a Z test taking the form of

$$Z = \frac{p_T - p_B}{\sqrt{p_0(1-p_0)(2/n + 1/c_T + 1/c_B)}},$$

where p_T and p_B were the estimated allele frequencies in the top and bottom pools, respectively; $p_0 = (p_T + p_B)/2$ was the allele frequency under the null hypothesis of $p_T = p_B$; n represented the number of chromosomes pooled; and c_T and c_B were the respective sequencing depths in the top and bottom pools. The Z statistic was distributed as standard normal under the null hypothesis, and P values were obtained accordingly.

Mutant Analyses. Mutations in 18 candidate genes from the GWA analysis (*5-HT7*, *Ac76E*, *Btk29A*, *CcapR*, *CG9932*, *Fak56D*, *Fas2*, *Fur1*, *fz*, *knrl*, *kuz*, *Pde1c*, *pk*, *Pkc53E*, *ppk11*, *Sdc*, *Spn*, *W*) and 18 candidate genes from the extreme QTL mapping analysis (*A2bp1*, *beat-1a*, *beat-IIIc*, *CG11321*, *CG15144*, *CG31961*, *CG43693*, *CG6959*, *chinmo*, *Cpr92F*, *Ih*, *Ddr*, *elk*, *mspo*, *Pvf3*, *sima*, *sm*, tor) were selected for functional assessment. *P{MiET1}* mutants (39) and their coisogenic controls were measured for olfactory responses to 0.3% (vol/vol) benzaldehyde, with five replicates (n = 50 flies per replicate) per line and sex. Statistically significant differences in responses to benzaldehyde between mutants and their coisogenic controls were determined by two-tailed t tests. The exceptions were *Pde1c*, *CG9932*, *Ac76E*, and *Spn*, which shared a common coisogenic control, and for which significance was tested using an adjusted Dunnett's test.

ACKNOWLEDGMENTS. We thank Sruthipriya Sridhar for technical assistance and Dr. Michael M. Magwire for helpful discussions and assistance with the GWA analyses. This work was supported by National Institutes of Health Grants R01 GM59469 and R01 GM45146 (to R.R.H.A. and T.F.C.M.).

1. Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753.
2. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90(1):7–24.
3. Moore JH, Williams SM (2009) Epistasis and its implications for personal genetics. *Am J Hum Genet* 85(3):309–320.
4. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 109(4): 1193–1198.
5. Wang K, Li M, Hakonarson H (2010) Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 11(12):843–854.
6. Califano A, Butte AJ, Friend S, Ideker T, Schadt E (2012) Leveraging models of cell regulation and GWAS data in integrative network-based association studies. *Nat Genet* 44(8):841–847.
7. Jia P, et al.; International Schizophrenia Consortium (2012) Network-assisted investigation of combined causal signals from genome-wide association studies in schizophrenia. *PLOS Comput Biol* 8(7):e1002587.
8. Pandey A, et al. (2012) Epistasis network centrality analysis yields pathway replication across two GWAS cohorts for bipolar disorder. *Transcult Psychiatry* 2:e154.
9. Mackay TFC, et al. (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482(7384):173–178.
10. Venken KJ, Bellen HJ (2012) Genome-wide manipulations of *Drosophila melanogaster* with transposons, Flp recombinase, and ΦC31 integrase. *Methods Mol Biol* 859: 203–228.
11. Huang W, et al. (2012) Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proc Natl Acad Sci USA* 109(39):15553–15559.
12. Su CY, Menuz K, Carlson JR (2009) Olfactory perception: Receptors, cells, and circuits. *Cell* 139(1):45–59.
13. Clyne PJ, et al. (1999) A novel family of divergent seven-transmembrane proteins: Candidate odorant receptors in Drosophila. *Neuron* 22(2):327–338.
14. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R (1999) A spatial map of olfactory receptor expression in the Drosophila antenna. *Cell* 96(5):725–736.
15. Gao Q, Chess A (1999) Identification of candidate Drosophila olfactory receptors from genomic DNA sequence. *Genomics* 60(1):31–39.
16. Hallem EA, Carlson JR (2006) Coding of odors by a receptor repertoire. *Cell* 125(1): 143–160.
17. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB (2009) Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell* 136(1):149–162.
18. Clyne PJ, Warr CG, Carlson JR (2000) Candidate taste receptors in Drosophila. *Science* 287(5459):1830–1834.
19. Scott K, et al. (2001) A chemosensory gene family encoding candidate gustatory and olfactory receptors in Drosophila. *Cell* 104(5):661–673.
20. Hekmat-Scafe DS, Scafe CR, McKinney AJ, Tanouye MA (2002) Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome Res* 12(9):1357–1369.
21. Galindo K, Smith DP (2001) A large family of divergent Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. *Genetics* 159(3):1059–1072.
22. Anholt RRH, Lyman RF, Mackay TFC (1996) Effects of single *P*-element insertions on olfactory behavior in *Drosophila melanogaster*. *Genetics* 143(1):293–301.
23. Sambandan D, Yamamoto A, Fanara JJ, Mackay TFC, Anholt RRH (2006) Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics* 174(3):1349–1363.
24. Fedorowicz GM, Fry JD, Anholt RRH, Mackay TFC (1998) Epistatic interactions between *smell-impaired* loci in *Drosophila melanogaster*. *Genetics* 148(4):1885–1891.
25. Mackay TFC, Hackett JB, Lyman RF, Wayne ML, Anholt RRH (1996) Quantitative genetic variation of odor-guided behavior in a natural population of *Drosophila melanogaster*. *Genetics* 144(2):727–735.
26. Wang P, Lyman RF, Shabalina SA, Mackay TFC, Anholt RRH (2007) Association of polymorphisms in odorant-binding protein genes with variation in olfactory response to benzaldehyde in Drosophila. *Genetics* 177(3):1655–1665.
27. Wang P, Lyman RF, Mackay TFC, Anholt RRH (2010) Natural variation in odorant recognition among odorant-binding proteins in *Drosophila melanogaster*. *Genetics* 184(3):759–767.
28. Arya GH, et al. (2010) Natural variation, functional pleiotropy and transcriptional contexts of odorant binding protein genes in *Drosophila melanogaster*. *Genetics* 186 (4):1475–1485.
29. Rollmann SM, et al. (2010) Odorant receptor polymorphisms and natural variation in olfactory behavior in *Drosophila melanogaster*. *Genetics* 186(2):687–697.
30. Ehrenreich IM, et al. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464(7291):1039–1042.
31. Paré G, Cook NR, Ridker PM, Chasman DI (2010) On the use of variance per genotype as a tool to identify quantitative trait interaction effects: A report from the Women's Genome Health Study. *PLoS Genet* 6(6):e1000981.
32. Struchalin MV, Dehghan A, Witteman JC, van Duijn C, Aulchenko YS (2010) Variance heterogeneity analysis for detection of potentially interacting genetic loci: Method and its limitations. *BMC Genet* 11:92.
33. Swarup S, et al. (2012) Extensive epistasis for olfactory behaviour, sleep and waking activity in *Drosophila melanogaster*. *Genet Res (Camb)* 94(1):9–20.
34. Swarup S, Williams TI, Anholt RRH (2011) Functional dissection of Odorant binding protein genes in *Drosophila melanogaster*. *Genes Brain Behav* 10(6):648–657.
35. Weber AL, et al. (2012) Genome-wide association analysis of oxidative stress resistance in *Drosophila melanogaster*. *PLoS ONE* 7(4):e34745.
36. Jordan KW, et al. (2012) Genome-wide association for sensitivity to chronic oxidative stress in *Drosophila melanogaster*. *PLoS ONE* 7(6):e38722.
37. Liu L, et al. (2003) Contribution of Drosophila DEG/ENaC genes to salt taste. *Neuron* 39(1):133–146.
38. Edwards AC, Zwarts L, Yamamoto A, Callaerts P, Mackay TFC (2009) Mutations in many genes affect aggressive behavior in *Drosophila melanogaster*. *BMC Biol* 7:29.
39. Bellen HJ, et al. (2011) The Drosophila gene disruption project: Progress using transposons with distinctive site specificities. *Genetics* 188(3):731–743.
40. Antonov AV, Schmidt EE, Dietmann S, Krestyaninova M, Hermjakob H (2010) R spider: A network-based analysis of gene lists by combining signaling and metabolic pathways from Reactome and KEGG databases. *Nucleic Acids Res* 38(Web Server issue): W78-83.
41. Koizumi K, et al. (2007) RNA interference screen to identify genes required for Drosophila embryonic nervous system development. *Proc Natl Acad Sci USA* 104(13): 5626–5631.
42. Pipes GC, Lin Q, Riley SE, Goodman CS (2001) The Beat generation: A multigene family encoding IgSF proteins related to the Beat axon guidance molecule in Drosophila. *Development* 128(22):4545–4552.
43. Moehring AJ, Mackay TFC (2004) The quantitative genetic basis of male mating behavior in *Drosophila melanogaster*. *Genetics* 167(3):1249–1263.
44. Zhu S, et al. (2006) Gradients of the Drosophila Chinmo BTB-zinc finger protein govern neuronal temporal identity. *Cell* 127(2):409–422.
45. Dekanty A, Lavista-Llanos S, Irisarri M, Oldham S, Wappner P (2005) The insulin-PI3K/TOR pathway induces a HIF-dependent transcriptional response in Drosophila by promoting nuclear localization of HIF-alpha/Sima. *J Cell Sci* 118(Pt 23):5431–5441.
46. Levene H (1960) *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, eds Olkin I, et al. (Stanford Univ Press, Stanford, CA), pp 278–292.
47. Wang Y, et al. (2001) Genetic manipulation of the odor-evoked distributed neural activity in the Drosophila mushroom body. *Neuron* 29(1):267–276.
48. Honegger KS, Campbell RA, Turner GC (2011) Cellular-resolution population imaging reveals robust sparse coding in the Drosophila mushroom body. *J Neurosci* 31(33): 11772–11785.
49. Xia S, Tully T (2007) Segregation of odor identity and intensity during odor discrimination in Drosophila mushroom body. *PLoS Biol* 5(10):e264.
50. Wong AM, Wang JW, Axel R (2002) Spatial representation of the glomerular map in the Drosophila protocerebrum. *Cell* 109(2):229–241.
51. Kauffman S (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ Press, Oxford, UK).
52. Greene CS, Penrod NM, Williams SM, Moore JH (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE* 4(6): e5639.
53. McQuilton P, St Pierre SE, Thurmond J; FlyBase Consortium (2012) FlyBase 101—the basics of navigating FlyBase. *Nucleic Acids Res* 40(Database issue):D706–D714.
54. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.