Accelerating gene discovery by phenotyping whole-genome sequenced multi-mutation strains and using the sequence kernel association test (SKAT)

Tiffany A. Timbers[1], Stephanie J. Garland[2], Swetha Mohan[1], Stephane Flibotte[2], Mark Edgley[2], Quintin Muncaster[1], Donald G. Moerman[2] and Michel R. Leroux[1,§]

[1]Department of Molecular Biology and Biochemistry and Centre for Cell Biology, Development, and Disease, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6 Canada

[2]Department of Zoology, University of British Columbia, 2350 Health Science Mall, Vancouver, British Columbia, V6T 1Z3 Canada

[§]Correspondence should be addressed to: leroux@sfu.ca

Forward genetic screens represent powerful, unbiased approaches to uncover novel components in any biological process. Such screens suffer from a major bottleneck, however, namely the cloning of corresponding genes causing the phenotypic variation. Reverse genetic screens have been employed as a way to circumvent this issue, but can often be limited in scope. Here we demonstrate an innovative approach to gene discovery. Using *C. elegans* as a model system, we used a whole-genome sequenced multi-mutation library together with the Sequence Kernel Association Test (SKAT) to rapidly screen for and identify genes associated with a phenotype of interest, namely defects in dye-filling of ciliated sensory neurons. Such anomalies in dye-filling are often associated with the disruption of cilia, organelles which in humans are implicated in sensory physiology (including vision, smell and hearing), development and disease. Beyond identifying several well characterised dye-filling genes, our approach uncovered 17 genes not previously linked to ciliated sensory neuron development or function. From these putative novel dye-filling genes, we confirmed and characterised the involvement of BGNT-1 in ciliated sensory neuron development. Notably, BGNT-1 is the orthologue of human B3GNT1/B4GAT1, a glycosyltransferase associated with Walker-Warburg syndrome, a multigenic disorder characterised by muscular dystrophy as well as brain and eye anomalies. Together, our work unveils an effective and innovative approach to gene discovery, and provides the first evidence that B3GNT1-associated Walker-Warburg syndrome is a ciliopathy.

Keywords: SKAT, GWAS, genetic screen, sensory biology, Million Mutation Project, cilia, *bgnt-1*, B3GNT1, B4GAT1, *C. elegans*

Introduction

A powerful, tried and true approach to identify which genes function in a particular biological process is to create collections of organisms harbouring multi-mutation variants *via* random mutagenesis, followed by screening the mutant library for organisms that exhibit the desired altered phenotypes. Although such forward genetics strategies have produced numerous fundamental discoveries, a significant limitation of this approach in metazoans is the prolonged time required to identify the causative mutations. The bottleneck typically arises from the required genetic mapping, complementation tests to exclude known genes, and sequencing of candidates genes.

To circumvent the major disadvantage of forward genetics, reverse genetic approaches have been employed. Various strategies for disrupting a collection of known genes (*e.g.,* RNAi, homologous recombination, transposon mutagenesis, *etc.*) are combined with phenotypic screening to identify candidates. Reverse genetics approaches also have drawbacks, however, including the need to handle and process tens of thousands of strains to assay the entire genome, off-target effects in the case of RNAi, and omission of essential genes.

We hypothesised that we could use advances in low-cost, whole-genome sequencing in combination with statistical genetics to inaugurate a novel gene discovery approach which retains the advantages of both forward and reverse genetics, yet minimises their downsides. To do this we employed the Million Mutation Project (MMP; Thompson et al. 2013), a collection of 2007 *Caenorhabditis elegans* strains harbouring randomly-induced mutations whose genomes are fully sequenced (data is publicly available: http://genome.sfu.ca/mmp/about.html). This mutant library represents an unprecedented genetic resource for any multicellular organism,

wherein the strains collectively contain one or more potentially disruptive alleles (~ 400 per strain) affecting nearly all *C. elegans* coding regions.

We postulated that this whole-genome sequence information would allow an "eyes wide open" approach when performing a genetic screen, such that pairing this resource with a high-throughput assay would enable rapid discovery of genes not previously associated with our biological process of interest. Here we demonstrate that testing for association between variants from the MMP library and phenotype data with the Sequence Kernel Association test (SKAT; Wu et al. 2011) allows us to effectively and efficiently predict novel genes important for our chosen biological process—the development and function of ciliated sensory neurons.

Results

Primary (non-motile) cilia arise from a modified centriole (basal body) and act as 'cellular antennae' that transduce environmental cues to the cell (Sung and Leroux 2013). They enable sensory physiology (such as olfaction/chemosensation, mechanosensation, vision) and are central to signalling pathways essential for metazoan development (Veland et al. 2009). Dysfunction of cilia is implicated in a number of human diseases, including polycystic kidney disease, congenital heart disease, and an emerging group of genetic disorders termed ciliopathies (*e.g.*, Bardet-Biedl, Meckel-Gruber and Joubert Syndromes). In these ciliopathies, disruption of many, if not all, cilia in the human body results in a plethora of defects, including retinal degeneration, organ cyst formation, obesity, brain malformations, and various other ailments (Baker and Beales 2009; Hildebrandt et al. 2011).

In *C. elegans*, the uptake of a fluorescent lipophilic dye, DiI, from the environment is used to probe the integrity of cilia and ciliated sensory neurons. DiI is selectively incorporated into six pairs of ciliated amphid channel sensory neurons in the head (ADF, ADL, ASH, ASI, ASJ, and ASK), and two pairs of ciliated phasmid channel sensory neurons in the tail (PHA and PHB), *via* environmentally-exposed cilia present at the tips of dendrites (**Suppl. Fig. 1**) (Perkins et al. 1986; Starich et al. 1995). Many dye-filling (*dyf*) mutants known from genetic screens (Hedgecock et al. 1985; Starich et al. 1995) harbour mutations in genes influencing ciliated sensory neuron development and function, including ciliogenesis (Inglis et al. 2007), cilia maintenance (Mohan et al. 2013), axon guidance (Hedgecock et al. 1985), dendrite anchoring/ formation (Heiman and Shaham 2009), cell fate (Herman and Horvitz 1994), and neural support (glial) cells (Bacaj et al. 2008).

We screened 480 randomly-chosen whole-genome sequenced multi-mutation MMP strains, ~25% of the library, for defects in DiI uptake in amphid and phasmid ciliated sensory neurons (**Fig. 1**). We found 40 MMP strains which exhibit significant amphid dye-filling defects and 40 MMP strains which exhibit significant phasmid dye-filling defects; notably, the strains exhibiting amphid and phasmid dye-filling defects are not necessarily identical (**Fig. 1c, Table 1, Suppl. Table 1**).

We identified 11 completely dye-fill defective strains, where all worms sampled failed to exhibit dye-filling. Of these, 10 contained deleterious ("knockout") mutations in previously identified dye-filling genes (*e.g.*, nonsense and frameshift-inducing deletions; **Suppl. Table 2**). Additionally, we uncovered 47 partially dye-fill defective strains, where a proportion of the

population failed to fill with dye significantly more often than wild-type worms. Of these partially dye-fill defective strains, 1 harbours a nonsense mutation and 10 display missense mutations in known dye-filling genes (**Suppl. Table 2**). In all, 37 dye-filling defective strains isolated are ostensibly caused by mutations in novel genes. Thus, having immediate access to genome sequences facilitates screen validation and identification of known dye-filling mutants / phenotypes ('positive' controls), as well as prioritisation of which strains to pursue further, based on the absence of mutations in known dye-filling genes.

To facilitate identification of potential novel genes responsible for the observed dye-fill defects, we hypothesised that a recently developed statistical genetics approach commonly used in human genetics, but underutilised in model organisms, would allow for the rapid prioritisation of candidate genes. Specifically we chose to employ the sequence kernel association test (SKAT) to identify genes associated with the dye-filling phenotype. SKAT is a regression method to test for association between rare and/or common genetic variants in a region and a continuous or dichotomous trait (Wu et al. 2011).

We chose SKAT over other statistical analyses for several reasons. For our dataset, it was imperative that we chose an association test that effectively dealt with rare variants, as 800,000/850,000 of the non-synonymous variants in the MMP library are rare; meaning that they are present in only a single isogenic strain in the library. Hence, genome-wide association study (GWAS) approaches, which typically test for an association between common variants (generally defined as a minor allele frequency > 5%) and a trait of interest, would be unsuitable for analysis of phenotype datasets derived from the MMP library. We also viewed SKAT as an optimal

method to use for our dataset because it permits the use of prior information to assign weights to genetic variants. For example, nonsense mutations might be expected to be more deleterious than other variants which may cause more modest changes to the protein, such as missense mutations and in-frame deletions. The C-alpha test (Neale et al. 2011), which is quite similar to SKAT in the absence of covariants (*e.g.,* age, sex, *etc.*), could have also been used for our dataset, but we chose to employ SKAT because it facilitates implementing and assigning biologically relevant weights to variants. Finally, SKAT was chosen over other related burden tests, such as the cohort allelic sums test (CAST; Morgenthaler and Thilly 2007) and the combined multivariate and collapsing (CMC) method (Li and Leal 2008), because unlike these tests, SKAT does not assume that all (common) variants will affect the trait in the same direction.

Given that the groups of worms which have amphid dye-filling and phasmid dye-filling defects do not necessarily overlap, we performed SKAT separately for each dataset. In its current implementation, SKAT can only accept dichotomous traits coded as 0 or 1 (it cannot input proportions), and thus for each dataset we coded strains whose neurons were significantly less likely to fill with dye compared to the wild-type control as 1 and all others as 0. Finally, we performed SKAT with biologically relevant weights assigned to the variants. We assigned mutations which would likely result in the creation of a null mutation (nonsense and splicing mutations, as well as frameshift causing deletions) a weight of 1, mutations which would result in truncation of the protein (in-frame deletions) a weight of 0.75 and mutations which would result in a change in amino acid sequence (missense mutation) a weight of 0.5. We hypothesised these were reasonable weights to assign to each class of mutation based on the current knowledge in the field of genetics.

Genome-wide SKAT analyses on the amphid dye-filling dataset revealed 18 genes that reached significance (Fisher's exact test) when we adjusted for multiple testing using a false discovery rate (FDR; Benjamini-Hochberg procedure) of 30% (**Table 2**, **Suppl. Table 3**). SKAT analyses for the phasmid dye-filling dataset uncovered 5 genes which reached significance, again using a FDR of 30% (**Table 3**, **Suppl. Table 4**). Dye-filling defects of both amphid and phasmid ciliated neurons was significantly associated with genes encoding intraflagellar transport proteins (OSM-1 and CHE-3), a glycosyltransferase (BGNT-1), as well as an Arf- and Rab-related protein (CNT-1; Table 2 & 3). Amphid-specific dye-filling defects were found to be associated with genes encoding a mitotic spindle assembly checkpoint protein, a tRNA synthetase, an ion channel regulator, a patched-domain containing protein, an aspartic protease, a protein-tyrosine phosphatase, a DMX-like protein, an inositol trisphosphate receptor, a JNK interacting protein, a GMP synthetase, an F-box A protein, a bidentate ribonuclease and a catenin delta protein (Table 2). Phasmid-specific dye-filling defects were found to be associated with a gene encoding the RAB6A GEF complex partner (Table 3).

Of the four genes associated with both amphid and phasmid dye-filling defects, namely *osm-1*, *che-3*, *cnt-1*, and *bgnt-1*, the first two are well characterised genes whose dye-filling defective phenotypes are ascribed to their key roles in intraflagellar transport (IFT). OSM-1 is the orthologue of mammalian IFT172, an IFT-B subcomplex component which functions as an adaptor to link ciliary cargo (*e.g.,* tubulin, receptors and signaling molecules) to the anterograde IFT kinesin motors, and is necessary for ciliogenesis (Signor et al. 1999). CHE-3, the orthologue of mammalian DYNC2H1, is a cytoplasmic dynein heavy chain which powers the retrograde

IFT-dynein motor. This molecular motor recycles IFT machinery from the growing ciliary tip back to the ciliary base (Signor et al. 1999; Wicks et al. 2000). These two known dye-fill/cilia genes represent excellent positive controls for our screen, and indicate that other genes found to be significantly associated with these phenotypes may be novel dye-fill genes that influence cilia function.

Interestingly, one of the other amphid/phasmid dye-filling gene hits, *cnt-1*, encodes a protein that plays a role in membrane trafficking/dynamics by influencing small GTPase function, *via* GTPase-activating protein or GAP activity. The general involvement of small GTPases of the Arf, Arf-like (Arl) and Rab families in cilium formation/development (Sung and Leroux 2013) is well established. *cnt-1* encodes the orthologue of human ACAP2, which interacts with both Rab35 (Kanno et al. 2010) and Arf6 (Jackson et al. 2000) to mediate crosstalk between these two proteins, at least in the context of PC12 cell neurite outgrowth, and potentially through endocytic recycling (Kobayashi and Fukuda 2012).

A third, putative novel dye-filling gene significantly associated with both amphid and phasmid dye-fill phenotypes is *bgnt-1* (**Table 2 & 3, Suppl. Table 3 & 4**). *bgnt-1* encodes an unstudied *C.elegans* glycosyltransferase 49 family member homologous to human B3GNT1/B4GAT1 (**Suppl. Fig 2**). B3GNT1 catalyses the addition of β1–3 linked N-acetylglucosamine to galactose (Zhou et al. 1999). In HeLa cells, its subcellular localisation is concentrated at the *trans*-Golgi (Lee et al. 2009). B3gnt1 knockout mice exhibit axon guidance phenotypes (Henion et al. 2005; Bless et al. 2006) and deficient behavioural responses to estrous females (Biellmann et al. 2008). In humans, mutations in B3GNT1 are associated with a congenital muscular dystrophy with

brain and eye anomalies, Walker-Warburg syndrome (WWS) (Buysse et al. 2013; Shaheen et al. 2013). WWS is a suspected, but unconfirmed ciliopathy; it exhibits 6 core features common to ciliopathies, including Dandy-Walker malformation, hypoplasia of the corpus callosum, mental retardation, posterior encephalocele, retinitis pigmentosa and *situs inversus* (Baker and Beales 2009). Additionally, one patient is reported to exhibit dysplastic kidneys (Buysse et al. 2013), a developmental disruption which leads to cyst formation, illuminating a potential 7th core ciliopathy feature to this disorder, renal cystic disease. To divulge a potential connection between B3GNT1 and cilia and/or ciliated sensory neuron function, we sought to confirm the role of *C. elegans* BGNT-1 in dye-filling, and analyse its involvement role in ciliated sensory neuron development.

Of the 8 MMP strains harbouring mutations in *bgnt-1*, VC20615, VC20628 and VC20326 exhibited severe dye-fill phenotypes (**Suppl. Table 1**). The C -> T missense mutation in *bgnt-1* in VC20615 corresponds to P194S alteration in the protein sequence, while VC20628 and VC20326 each harbour an identical G -> A missense mutation in *bgnt-1* which leads to a P194S amino acid change in the protein sequence. Both of these mutations alter conserved amino acid residues (**Suppl. Fig. 3**). To confirm that the mutations in *bgnt-1* was responsible for the dye-filling phenotypes in *bgnt-1* mutants we rescued the dye-fill defects by expressing a fosmid containing a wild-type copy of *bgnt-1* in an extrachromosomal array (**Fig. 2**). Another way to confirm that disruption of *bgnt-1* causes dye-fill defects would be to observe this phenotype in a strain harbouring a knock-out mutation in *bgnt-1*. Unfortunately, a knock-out allele of *bgnt-1* does not exist yet, and thus we tested the causality of *bgnt-1* via a relatively efficient SNP mapping approach. We established that the dye-fill phenotypes from VC20615 and VC20628

strains mapped to the *bgnt-1* locus, on chromosome IV between -5 cM and 8 cM (**Suppl. Fig. 4**). Notably, in both VC20615 and VC20628 strains, *bgnt-1* was the only gene in this region harbouring a mutation which was common to both of these strains. These findings indicate that the *bgnt-1* mutations in VC20615, VC20628 and VC20326 are causative for the observed dye-filling defects in these strains.

To shed light on how *bgnt-1* affects dye-filling, we expressed GFP-tagged BGNT-1, driven by its endogenous promoter, in *C. elegans*. We observed that the protein localises to the cytosol of ADL ciliated sensory head (amphid) neurons, and the glial-like phasmid sheath (PHsh) cells in the tail that are closely associated with the ciliated phasmid sensory neurons (**Fig. 3, Suppl. Fig. 5**). Next, we queried whether *bgnt-1* mutants exhibit any gross ciliary morphology defects by expressing a ciliary marker in *bgnt-1* mutants, namely the GFP-tagged IFT-B subcomplex protein, CHE-2 (IFT80). This experiment indicated that although the cilia of *bgnt-1* mutants fail to fill with dye, their ciliary structures appear superficially wild-type (**Fig. 4a**).

Given that *bgnt-1* is only expressed in the ADL amphid ciliated sensory neurons, we decided to further characterise the phenotype of ADL cilia and dendrites using the primarily cell-specific ADL promoter, P*srh-220*, to drive expression of another cilia marker, IFT-20 (IFT20) tagged with tdTomato, in wild-type and *bgnt-1* mutants. In this strain we also expressed cytoplasmic GFP in the amphid socket cells so that we could evaluate whether or not the ADL cilia were correctly associated with the surrounding glial support cells and the pore where DiI has access to the amphid ciliated sensory neurons from the environment. Similar to the experiment with the CHE-2::GFP pan-cilia marker, the P*srh-220*::IFT-20::tdTomato marker revealed that the ADL

cilia and amphid socket (Amso) cell morphology also appear superficially wild-type in *bgnt-1* mutants (**Fig. 4b**). We then sought to assay for potential phenotypes involving ADL cilia length (**Fig. 4c**); length of socket cell penetration by ADL (proxied by the distance from the distal tip of ADL cilia to the distal end of the socket cell tip; **Fig. 4d**) ADL guidance (proportion of double rod cilia/amphid; **Fig. 4e**); and finally, ADL dendrite blebbing (structural alteration where dendrites take bead on a string appearance; **Fig. 4f**). Our analyses revealed that ADL cilia in *bgnt-1* mutants are wild-type in virtually every aspect except for a modest cilia length defect. *bgnt-1* mutants were observed to have significantly longer cilia compared to wild-type worms (**Fig. 4c**; $p < 0.01$, Kruskal-Wallis test).

These results lead us to hypothesise that BGNT-1 plays a role in the development and/or function of the amphid and phasmid sensilla but does not influence gross cilia morphology. Although *bgnt-1* mutants were observed to have an increased cilium length phenotype, we might not predict that such a marginal increase in length results in abrogated dye filling. One possibility is that the glycosyltransferase regulates the association of cilia with the sheath and socket glial-like cells which envelop them (Perkins et al. 1986) (**Suppl. Fig. 1**). This defect may not be visible at the level of light microscopy and could perhaps result from changes to the lamellar membrane that surround the amphid/phasmid cilia or the secreted extracellular material lining these channels (Perkins et al. 1986). Which substrate(s) the β1,3-N-acetylglucosaminyltransferase, BGNT-1 (B3GNT1), glycosylates, and how this influences sensory neuron/glial cell development and function, remains to be determined.

Interestingly, BGNT-1 is only expressed in the ADL amphid ciliated sensory neurons, making the severe dye-fill phenotype in amphid neurons a puzzling observation (in addition to ADL, 5 other neuron pairs dye-fill). One hypothesis that could explain this is that not all 12 amphid channel cilia are equally important for amphid sensillum morphogenesis and that during ciliogenesis, one or more "keystone" cilia may be critical for establishing the correct lamellar membrane or composition of secreted extracellular material that protects the amphid pore. Evidence for such "keystone" cilia were obtained when dye-filling was restored by cell-specific genetic rescue of a *daf-19* mutant, which cannot form cilia and is thus dye-fill defective. Remarkably, expression of wild-type *daf-19* in some amphid ciliated sensory neurons, including ADL, but not others, restored dye-filling (Senti et al. 2009). Hence, BGNT-1 could be required for ADL cilia and/or sensory neuron development, a function critical for supporting ciliary access to the environment.

We attempted to test this hypothesis by creating a suite of transgenic worms harbouring the G205E mutation (*gk361915* allele) which also expressed wild-type *bgnt-1* under control of cell-specific promoters as an extrachromosomal array. Promoters tested included *bgnt-1* (ADL and the phasmid sheath cell expression), *bbs-8* (pan-cilia expression), *ver-1* (amphid and phasmid sheath cell expression) and a strain which expressed both the *bbs-8* and *ver-1* driven constructs. None of these transgenic lines were able to rescue the dye-filling defects of *bgnt-1* mutants (**Suppl. Fig. 6**), indicating that either the expression pattern of *bgnt-1* is wider than what we observed in this study, or *bgnt-1* rescue is sensitive to copy number and the high-level of over-expression generated by expressing transgenes *via* an extrachromosomal array leads to a dominant negative or gene silencing effect (e.g., tens to hundreds Fire and Waterston 1989; Praitis et al. 2001).

In humans, mutations in *B3GNT1* cause Walker-Warburg syndrome (WWS; Buysse et al. 2013; Shaheen et al. 2013). Given that mutations in B3GNT1 lead to WWS and that it is classified as a dystroglycanopathy, a group of muscular disorders whose etiology is hypothesised to be caused by aberrant glycosylation of dystroglycan, we tested whether or not the *C. elegans* dystroglycan homologs, *dgn-1, dgn-2* and *dgn-3*, exhibited dye-filling phenotypes. We observed that all *dgn* mutants exhibited dye-filling indistinguishable from wild-type worms (**Suppl. Fig. 7**), indicating that BGNT-1 function in dye-filling is likely independent of dystroglycan. Interestingly, as highlighted earlier, the WWS congenital muscular dystrophy exhibits 6 features beyond muscle structure/function disruption which are core ciliary disorder (ciliopathies) features (Baker and Beales 2009). Our findings that *C. elegans bgnt-1* is expressed in ciliated cells or associated glial cells, influences a cilium-dependent phenotype (dye-filling), and functions independently of dystroglycan supports the notion that WWS may be a novel ciliopathy. It also underscores the importance of identifying novel dye-filling genes, some of which might be implicated in human ciliopathies.

Discussion

Here we demonstrate that using genome-wide association analysis (*e.g.,* SKAT) is an efficient way to rapidly uncover novel genes for a phenotype of interest (*e.g.*, ciliated sensory neuron function) in whole-genome sequenced strains harbouring multiple mutations induced via random mutagenesis. We found that 3 cilia-related genes, osm-1, che-3 and bgnt-1 were significantly associated with dye-filling defects, suggesting that ~ 70% of the remaining 16 genes significantly associated with this phenotype likely represent genes important for ciliary/sensory neuron

development and/or function. We characterised *bgnt-1,* a gene SKAT identified as being associated with the dye-filling phenotypes but not previously implicated in cilia or amphid-sensillum function. We observed that: *(1)* two missense mutations in *bgnt-1* result in severe dye-fill defects in 3 MMP strains; *(2)* a fosmid containing full-length wild-type *bgnt-1* rescues the dye-filling phenotype in *bgnt-1* mutants; *(3)* the dye-filling phenotypes in the MMP strains with mutations in *bgnt-1* map to the *bgnt-1* locus; and finally, *(4)* BGTN-1 is expressed in the ADL ciliated sensory neuron and the phasmid sheath cell. Together, these data strongly indicate that BGNT-1 functions in dye-filling, but further phenotyping of *bgnt-1* mutant cilia and glial support cells, as well as cell-specific rescue experiments, were unable to indicate the cellular/molecular reasons for the dye-filling defect in *bgnt-1* mutants.

SKAT analysis of the 480 strains without weights resulted in a similar list of genes not previously known to be associated with the dye-filling phenotype, but did not result in known dye-filling genes, *osm-1* and *che-3*, being significantly associated with dye-filling defects (**Supp. Table 5 & 6**). We hypothesise that *osm-1* and *che-3* were not found to be significantly associated with the dye-filling phenotype in the analysis which weighted all variants equally because without weights, the effect of gene size (odds ratio) on dye-filling for these genes is small. This results from the many MMP missense mutation alleles of each of these genes that do not cause a dye-filling defect. Thus, to capture genes which cause dye-filling defects when they harbour missense mutations (*e.g.*, *bgnt-1*), as well as those which only cause dye-filling defects when they contain knockout mutations (*e.g., che-3* and *osm-1*), we recommend assigning biologically relevant weights when using SKAT with the MMP library. The weight assignment could be simple, as done here, or more complex, calculating, for example, the SIFT (Kumar et al. 2009) or

Polyphen (Ramensky et al. 2002) scores for assessing the severity of each variant in the MMP library.

The genome-wide statistical genetic approach presented here has several advantages over traditional screening approaches. It generates a prioritised list of candidate genes likely responsible for the phenotype of interest. After this list is generated *via* screening and SKAT analysis, candidates can be tested for their causality of the phenotype through several standard genetic approaches in *C. elegans*. Candidates could be confirmed, for example, by (i) testing for the phenotype in knock-out mutants or RNAi, (ii) genetic rescue experiments, (iii) performing a genetic complementation test between two loss of function alleles, or (iv) mapping the mutation to the gene locus. Thus, this strategy may work for phenotypes where the traditional polymorphic SNP-mapping strain, CB4856, diverges from the reference wild-type strain, N2, from which the MMP library was generated (Thompson et al. 2013), as well as partially-penetrant or other difficult-to-score phenotypes. In the case of *bgnt-1* we performed genetic rescue experiments and SNP mapping to support our the SKAT findings, which indicates that *bgnt-1* mutations cause dye-filling defects. This was done rather than more efficient and straight-forward approaches listed above because there is no knock-out allele for *bgnt-1*, and because *C. elegans* nervous system expressed genes are refractory to RNAi.

Another potential extension and utility of this approach that could work for some (non-neural) phenotypes would be pairing the screening of the MMP strains with RNAi to look for enhancing, suppressing or synthetic phenotypes, and then using SKAT to prioritise a list of candidate genes. Furthermore, as more data is collected on the MMP strains, data from multiple phenotypes could

be combined to perform multi-variate genome-wide statistical analysis on whole-genome sequence data. Such approaches have to been shown to be more powerful than univariate approaches in the case of SNP array data (e.g. Ferreira and Purcell 2009; O'Reilly et al. 2012; van der Sluis et al. 2013) and such tests can also indicate which variants are pleiotropic, or specific to a single phenotype. How to perform this multi-variate phenotype analysis on whole-genome sequences is currently an active area of research and tools to make this possible are being developed.

There are also challenges and limitations to the statistical genetic approach presented here. First, this approach of performing a "medium"-scale screen of the MMP strains is limited to assays that can be done without genetic manipulation of the strains. For example, introducing a transgene into 480 strains would require a prohibitive amount of work; although this has been done for 90 MMP strains (Wang et al. 2015). Second, the statistical analysis presented here is only possible for genes which have > 1 variant in the population of worms screened. In practice we found it works optimally for strains with at least 7 variants. This is due to the distribution of p-values when attempting to control for multiple testing; in our dataset, fewer than 7 variants led to a skewed p-value distribution and an inflation of q-values during False-discovery rate adjustment. This strict rule demanding high-coverage for our SKAT analysis leads to only 1150 genes in the 480 MMP strains being considered here. This is due to the distribution of minor allele counts in the MMP strains (**Suppl. Fig. 8**), which exponentially decreases from 1 to N. If the end goal is to simply obtain a prioritised list of genes to follow-up on from a screen, controlling for multiple testing issues may not be necessary as it does not change the rank of genes on the list. If this

approach is taken, we would advise that only genes with a minimum minor allele count of > 3 be used.

Finally, the power of genome-wide association analysis augments as the number of strains increases (the probability of additional mutations in specific genes is increased), and thus, screening the entire MMP library would likely uncover many additional genes associated with dye-filling defects. We can estimate that using the SKAT approach with the biologically relevant weights, as done here, one would need to screen a minimum of ~40-160 MMP strains to find at least one gene associated with a phenotype of interest. For example, a false-discovery rate of 30% allowed us to uncover 5 genes exhibiting a phasmid dye-filling phenotype, meaning that 3 of these genes are likely truly associated with the phenotype. This was uncovered using 480 strains, thus to find a single gene associated with this phenotype we would need to screen ~160 strains (480/3). Using similar logic for amphids, we would need to screen ~40 strains. This empirical estimation relies on the prevalence of phenotypes being similar to those we observed for amphid and phasmid dye-filling (8% of MMP strains). Our unpublished data from analysing ~500 MMP strains revealed a similar prevalence (~10%) for the muscle disorganisation phenotype, as assessed by polarised light microscopy; hence, it is not unreasonable that this may also be observed for other phenotypes.

In conclusion, we demonstrated the utility and efficiency of using deep-sequenced multi-mutant strains in combination with SKAT to rapidly uncover novel genes required for a biological process of interest—here, ciliated sensory neuron development and/or function. For all new putative dye-filling genes highlighted in this study, we had no prior knowledge of their

importance in ciliated sensory neuron function, and may not have (easily) uncovered them using alternative methods. Our approach therefore reduces the hurdle of traditional forward genetic methods, namely identifying the causative allele, and improves upon reverse genetics by allowing high gene/mutation coverage in a relatively small number of strains. Lastly, we propose that our approach is applicable not only for *C. elegans*, but any organism with a small genome that can be quickly sequenced and where numerous mutant strains can be isolated and phenotyped with relative ease, including *Drosophila* and *Arabidopsis.*

Methods

Strains and maintenance. Worms were cultured on Nematode Growth Medium (NGM) seeded with Escherichia coli (OP50) at 20ºC as described previously (Brenner 1974). The following strains were obtained from the Caenorhabditis Genetics Center (University of Minnesota, Minneapolis, MN): N2 Bristol, CB4856, CH1869, CH1878 and PR813. VC2010, the wild-type reference strain used during the dye-filling screen, was derived from N2 (Thompson et al. 2013). The Million Mutation Project strains were isolated and their genomes' sequenced by Thompson *et al*. (2013). The 480 Million Mutation Project strains used in this study are listed in **Table S1**.

Dye-filling procedures.

Dye-filling assays were performed using the fluorescent dye DiI (Molecular Probes; DiIC18 Vybrant DiI cell-labelling solution, diluted 1:1000 with M9 buffer). Mixed stage *C. elegans* cultures were stained for 30 minutes, and Dil uptake into the amphid and phasmid neurons was visualised using either a Zeiss fluorescent dissection scope (dye-filling screen) or spinning disc confocal microscope (WaveFX spinning disc confocal system from Quorum Technologies) using a 25X oil (N.A 0.8) objective and Hammamatsu 9100 EMCCD camera. Volocity software (PerkinElmer) was used for acquisition. The completely dye-filling defective (*dyf*) mutant strain PR813 *osm-5(p813)* was used as a positive control for the dye-filling phenotype.

For the dye-filling screen, two plates of mixed-stage *C. elegans* were dye-filled for each Million Mutation Project strain, and defects were quantified by counting the number of worms exhibiting amphid and/or phasmid dye-filling defects. A worm was classified to have a dye-filling defect if: *i)* no fluorescence was observed, *ii)* fluorescence was observed to be greatly reduced (minimum

of ~ estimated ~3x fluorescence reduction compared to wild-type staining from the experiment at the same magnification and laser intensity) and/or *iii)* fluorescence staining pattern was abrogated (e.g. accumulations of fluorescence at tips of dendrites with little to no staining in cell bodies). Fifteen worms were scored from each plate. If the dye-filling of a Million Mutation Project strain appeared qualitatively dimmer than wild-type worms across both plates or if ≥ 25% of the population exhibited a dye-filling defect the assay was repeated for that strain. A Fisher's exact test followed by p-value adjustment using false discovery rate of 5% (Benjamini–Hochberg procedure) was used to assign strains a value of 1 if they exhibited a significant dye-fill defect compared to wild-type (N2), or 0 if indistinguishable from wild-type. This was done separately for both amphids and phasmids.

Identification, mapping and cloning of bgnt-1.

To rough-map the dye-filling defects of Million Mutation Project strains to an arm of a chromosome we used the high-throughput SNP mapping approach created by Davis *et al*. (2005). The following SNPs used by Davis *et al*. (2005) were omitted from our analysis because the whole genome sequence data from Thompson *et al.* (2013) could not safely deduce that the SNPs from parental strain subjected to mutagenesis, VC2010 (from which the Million Mutation Project strains were generated), matched those of Bristol N2 but not Hawaiin CB4856 (mapping strain): W03D8, F58D5, T01D1, Y6D1A, Y38E10A, T12B5, R10D12, F11A1, and T24C2.

Preparation of transgenic lines.

For native rescue of VC20628 *bgnt-1(gk361915)*, 25 ng/µl of fosmid WRM065bB05 containing *bgnt-1* was injected into *bgnt-1* mutants along with 80 ng/µl of pRF4 *rol-6(su1006dm)* as a co-

injection marker. *bgnt-1(gk361915)*; Ex[CHE-2::GFP; pRF4] was created by crossing *bgnt-1(gk361915)* with wild-type worms expressing Ex[CHE-2::GFP; pRF4]. The translational Psrh-220::IFT-20::tdTomato fusion was generated as described in (Mohan et al. 2013), except that tdTomato was used in place of GFP. 1 µl of the PCR product was microinjected into germline of gravid worms along with a co-injection markers (pRF4 *rol-6*(*su1006dm*), final concentration of 100 ng/µl). Stable lines expressing this extrachromosomal array were crossed into DM13283 *dpy-5(e907)*; sIs12964[P*grd-15*::GFP; pCeh361] to create the strain MX1924 *dpy-5(e907)*; Ex[P*srh-220*::IFT-20::tdTomato; pRF4]; sIs12964[P*grd-15*::GFP; pCeh361]. *bgnt-1(gk361915)* was also introduced to this line via genetic crossing to create MX2236 *bgnt-1(gk361915); dpy-5(e907)*; Ex[P*srh-220*::IFT-20::tdTomato; pRF4]; sIs12964[P*grd-15*::GFP; pCeh361]. The translational BGNT-1::GFP fusion construct was generated by using PCR to fuse GFP::*unc-54* 3' UTR (amplified from the GFP expression vector pPD95.77) to the promoter (3346 bp) and genomic DNA of the *bgnt-1* gene, removing the *bgnt-1* stop codon. 1 µl of the PCR product was microinjected into germline of gravid worms along with a co-injection markers (pRF4 *rol-6*(*su1006dm*), final concentration of 80 ng/µl), and the ciliary marker (P*osm-5*::XBX-1::tdTomato, final concentration of 20 ng/µl). The P*bbs-8*::BGNT-1 cell specific rescue construct was created by PCR fusion to stitch 841 bp of the *bbs-8* promoter to the bgnt-1 cDNA and 920 bp of the *unc-54* 3' UTR. The P*ver-1*::BGNT-1 cell-specific rescue construct was created by using PCR fusion to stitch 1982 bp of the *ver-1* promoter to the bgnt-1 cDNA and then blunt cloning this into pJET1.2 (selecting for the reverse orientation). The 1021 bp of the *unc-54* 3' UTR was then amplified via PCR with NcoI adapters. Both P*ver-1*::BGNT-1 in pJET1.2 and the *unc-54* 3' UTR PCR product were digested with NcoI and the *unc-54* 3' UTR PCR product was ligated into the pJET1.2 plasmid containing P*ver-1*::BGNT-1. Digestion was

used to select for plasmids with the *unc-54* 3' UTR in the correct orientation and was confirmed via sequencing. 10 ng/ul of the rescue construct was microinjected into the germline of gravid worms along with the coinjection marker, pRF4 *rol-6(su1006dm)*, at a concentration of 90 ng/ul.

Imaging sensory neurons and cilia.

For visualisation of fluorescent-tagged proteins, worms were immobilised in 1µl of 25mM levamisole and 1µl of 0.1µm diameter polystyrene microspheres (Polysciences 00876-15, 2.5% w/v suspension) on 10% agarose pads and visualised under a spinning disc confocal microscope (WaveFX spinning disc confocal system from Quorum Technologies) using a 100X oil (N.A 1.4) objective and Hammamatsu 9100 EMCCD camera. Volocity 6.3 was used to deconvolve images as well as measure ADL cilia length and distal tip of ADL cilia to distal end of amphid socket cell length. The researcher was blind while performing the quantisation of ADL cilia/dendrite phenotypes.

Phylogenetic Analysis

Protein sequences (obtained from: http://www.cazy.org/) were aligned using MUSCLE 3.7 (Edgar, 2004). The phylogenetic tree was built using PhyML 3.0 aLRT (Guindon *et al.*, 2010) and viewed using FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

SKAT analysis

We performed SKAT using the SKAT package (version 1.0.9) (Wu et al. 2011) in R (version 3.2.1) with the RStudio (version 0.99.467) developing environment. No covariates were used. Given that the MMP library was created *via* random mutagenesis of the same isogenic parental

strain (Thompson et al. 2013), and thus we did not have to control for population stratification. Custom, biologically relevant weights were assigned to the variants. Nonsense, splicing mutations and frameshift causing deletions were assigned a weight of 1, in-frame deletions were assigned a weight of 0.75, and missense mutations were assigned a weight of 0.25. Gene-based tests for all genes with a minor allele count > 6 were performed. A false discovery rate (Benjamini-Hochberg procedure) of 30% was used to determine genes which were significantly associated with the phenotype. Make, Perl and R scripts used to perform the analysis can be found at: https://github.com/ttimbers/Million-Mutation-Project-dye-filling-SKAT.git

Acknowledgements

M.R.L. acknowledges funding from the Canadian Institutes of Health Research (CIHR; MOP82870) and a senior scholar award from Michael Smith Foundation for Health Research (MSFHR). The work in the D.G.M. laboratory was supported by CIHR. D.G.M. is a Senior Fellow of the Canadian Institute For Advanced Research. S.J.G. acknowledges funding from a Natural Sciences and Engineering Research Council (NSERC) Undergraduate Student Research Award (USRA).

Author Contributions

T.A.T., M.R.L., and D.G.M. developed the methodology of how to screen and identify novel genes from the MMP library. T.A.T., and S.J.G. performed the dye-fill screen of the library. M.E. assisted with the screen by maintaining and organizing mutant strains. T.A.T. performed the

genomic and statistical analysis, as well as wrote all Make and R scripts. S.F. advised T.A.T. on genomic analysis and wrote the perl script to filter .vcf files. T.A.T. performed SNP-mapping, and all other experiments related to *bgnt-1*. S.M. provided intellectual input on and assisted with molecular biology and imaging experiments. T.A.T. prepared the figures, and wrote the manuscript. All authors read and commented on the manuscript.

References

Bacaj T, Tevlin M, Lu Y, Shaham S. 2008. Glia are essential for sensory organ function in C. elegans. Science 322: 744-747.

Baker K, Beales PL. 2009. Making sense of cilia in disease: the human ciliopathies. Am J Med Genet C Semin Med Genet 151C: 281-295.

Biellmann F, Henion TR, Burki K, Hennet T. 2008. Impaired sexual behavior in male mice deficient for the beta1-3 N-acetylglucosaminyltransferase-I gene. Molecular reproduction and development 75: 699-706.

Bless E, Raitcheva D, Henion TR, Tobet S, Schwarting GA. 2006. Lactosamine modulates the rate of migration of GnRH neurons during mouse development. The European journal of neuroscience 24: 654-660.

Brenner S. 1974. The genetics of Caenorhabditis elegans. Genetics 77: 71-94.

Buysse K, Riemersma M, Powell G, van Reeuwijk J, Chitayat D, Roscioli T, Kamsteeg EJ, van den Elzen C, van Beusekom E, Blaser S et al. 2013. Missense mutations in beta-1,3-N-acetylglucosaminyltransferase 1 (B3GNT1) cause Walker-Warburg syndrome. Human molecular genetics 22: 1746-1754.

Davis MW, Hammarlund M, Harrach T, Hullett P, Olsen S, Jorgensen EM. 2005. Rapid single nucleotide polymorphism mapping in C. elegans. BMC Genomics 6: 118.

Ferreira MA, Purcell SM. 2009. A multivariate test of association. Bioinformatics 25: 132-133.

Fire A, Waterston RH. 1989. Proper expression of myosin genes in transgenic nematodes. EMBO J 8: 3419-3428.

Hedgecock EM, Culotti JG, Thomson JN, Perkins LA. 1985. Axonal guidance mutants of Caenorhabditis elegans identified by filling sensory neurons with fluorescein dyes. Dev Biol 111: 158-170.

Heiman MG, Shaham S. 2009. DEX-1 and DYF-7 establish sensory dendrite length by anchoring dendritic tips during cell migration. Cell 137: 344-355.

Henion TR, Raitcheva D, Grosholz R, Biellmann F, Skarnes WC, Hennet T, Schwarting GA. 2005. Beta1,3-N-acetylglucosaminyltransferase 1 glycosylation is required for axon pathfinding by olfactory sensory neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 25: 1894-1903.

Herman MA, Horvitz HR. 1994. The Caenorhabditis elegans gene lin-44 controls the polarity of asymmetric cell divisions. Development 120: 1035-1047.

Hildebrandt F, Benzing T, Katsanis N. 2011. Ciliopathies. N Engl J Med 364: 1533-1543.

Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A et al. 2007. High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. PLoS Biol 5: e237.

Inglis PN, Ou G, Leroux MR, Scholey JM. 2007. The sensory cilia of Caenorhabditis elegans. WormBook doi:10.1895/wormbook.1.126.2: 1-22.

Jackson TR, Brown FD, Nie Z, Miura K, Foroni L, Sun J, Hsu VW, Donaldson JG, Randazzo PA. 2000. ACAPs are arf6 GTPase-activating proteins that function in the cell periphery. The Journal of cell biology 151: 627-638.

Kanno E, Ishibashi K, Kobayashi H, Matsui T, Ohbayashi N, Fukuda M. 2010. Comprehensive screening for novel rab-binding proteins by GST pull-down assay using 60 different mammalian Rabs. Traffic 11: 491-507.

Kobayashi H, Fukuda M. 2012. Rab35 regulates Arf6 activity through centaurin-beta2 (ACAP2) during neurite outgrowth. Journal of cell science 125: 2235-2243.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073-1081.

Lee PL, Kohler JJ, Pfeffer SR. 2009. Association of beta-1,3-N-acetylglucosaminyltransferase 1 and beta-1,4-galactosyltransferase 1, trans-Golgi enzymes involved in coupled poly-N-acetyllactosamine synthesis. Glycobiology 19: 655-664.

Li B, Leal SM. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. American journal of human genetics 83: 311-321.

Mohan S, Timbers TA, Kennedy J, Blacque OE, Leroux MR. 2013. Striated rootlet and non-filamentous forms of rootletin maintain ciliary function. Current Biology 23: 2016-2022.

Morgenthaler S, Thilly WG. 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutation research 615: 28-56.

Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M, Kathiresan S, Purcell SM, Roeder K, Daly MJ. 2011. Testing for an unusual distribution of rare variants. PLoS genetics 7: e1001322.

O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin MR, Coin LJ. 2012. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS One 7: e34861.

Perkins LA, Hedgecock EM, Thomson JN, Culotti JG. 1986. Mutant sensory cilia in the nematode Caenorhabditis elegans. Dev Biol 117: 456-487.

Praitis V, Casey E, Collar D, Austin J. 2001. Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics 157: 1217-1226.

Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30: 3894-3900.

Senti G, Ezcurra M, Lobner J, Schafer WR, Swoboda P. 2009. Worms with a single functional sensory cilium generate proper neuron-specific behavioral output. Genetics 183: 595-605, 591SI-593SI.

Shaheen R, Faqeih E, Ansari S, Alkuraya FS. 2013. A truncating mutation in B3GNT1 causes severe Walker-Warburg syndrome. Neurogenetics doi:10.1007/s10048-013-0367-8.

Signor D, Wedaman KP, Orozco JT, Dwyer ND, Bargmann CI, Rose LS, Scholey JM. 1999. Role of a class DHC1b dynein in retrograde transport of IFT motors and IFT raft particles along cilia, but not dendrites, in chemosensory neurons of living Caenorhabditis elegans. The Journal of cell biology 147: 519-530.

Starich TA, Herman RK, Kari CK, Yeh WH, Schackwitz WS, Schuyler MW, Collet J, Thomas JH, Riddle DL. 1995. Mutations affecting the chemosensory neurons of Caenorhabditis elegans. Genetics 139: 171-188.

Sung CH, Leroux MR. 2013. The roles of evolutionarily conserved functional modules in cilia-related trafficking. Nature cell biology 15: 1387-1397.

Thompson O, Edgley M, Strasbourger P, Flibotte S, Ewing B, Adair R, Au V, Chaudry I, Fernando L, Hutter H et al. 2013. The Million Mutation Project: A new approach to genetics in Caenorhabditis elegans. Genome Res doi:gr.157651.113 [pii] 10.1101/gr. 157651.113.

van der Sluis S, Posthuma D, Dolan CV. 2013. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genet 9: e1003235.

Veland IR, Awan A, Pedersen LB, Yoder BK, Christensen ST. 2009. Primary cilia and signaling pathways in mammalian development, health and disease. Nephron Physiol 111: p39-53.

Wang X, Liu J, Zhu Z, Ou G. 2015. The heparan sulfate-modifying enzyme glucuronyl C5-epimerase HSE-5 controls Caenorhabditis elegans Q neuroblast polarization during migration. Dev Biol 399: 306-314.

Wicks SR, de Vries CJ, van Luenen HG, Plasterk RH. 2000. CHE-3, a cytosolic dynein heavy chain, is required for sensory cilia structure and function in Caenorhabditis elegans. Developmental biology 221: 295-307.

Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89: 82-93.

Zhou D, Berger EG, Hennet T. 1999. Molecular cloning of a human UDP-galactose:GlcNAcbeta1,3GalNAc beta1, 3 galactosyltransferase gene encoding an O-linked core3-elongation enzyme. European journal of biochemistry / FEBS 263: 571-576.

Figure Captions

Figure 1. Dye-filling (ciliated sensory neuron development/function) screening methodology and results.

(a,b) Input to the screen was 480 whole genome-sequenced multi-mutant strains from the Million Mutation Project (Thompson et al. 2013). Mixed-stage *C. elegans* cultures were incubated with DiI for 30 minutes, washed in buffer and then examined by fluorescence microscopy for their ability to uptake the dye into head (amphid) and tail (phasmid) sensory neurons.

(c) Dye-filling phenotypes of each of the 480 MMP strains which were assayed. The proportion of worms exhibiting dye-filling in strains represented by dark grey diamonds were not statistically separable from the proportion of wild-type worms exhibiting dye-filling as assessed by a Fisher's exact test with p-values adjusted for a 5% false discovery rate (Benjamini–Hochberg procedure) to control for multiple testing. Blue and red diamonds represent strains with mainly or exclusively amphid or phasmid dye-filling defects, while purple diamonds show strains with defects in both sensory neurons. Two highlighted strains, VC20615 and VC20628, contain mutations which alter conserved amino acid residues in the protein encoded by *C. elegans bgnt-1,* a gene identified by SKAT to be associated with both amphid and phasmid dye-filling defects.

Figure 2. Cloning of *bgnt-1* as a novel dye-filling gene.

(a) Wild-type *bgnt-1* rescues dye-filling defects in *bgnt-1 (gk361915)* mutants. Transformation of VC20628 *bgnt-1 (gk361915)* with a fosmid containing wild-type *bgnt-1* (WRM065bB05)

completely rescues the amphid, and partially rescues the phasmid dye-filling defects of VC20628 *bgnt-1 (gk361915)* mutants, as assessed by fluorescence microscopy.

(b) Quantitation of amphid and phasmid dye-filling in the mutant strain VC20628 in the presence or absence of a fosmid rescue contruct ($p < 0.05$, Fisher's exact test). Error bars represent 95% confidence intervals (Pearson Clopper method).

Figure 3. *C. elegans* BGNT-1 localises to ADL ciliated sensory neurons and phasmid sheath cell. A translational C-terminal fusion of GFP with the endogenous *bgnt-1* promoter and coding region localises diffusely to a pair of ciliated sensory neurons in the head (ADL left and right), and a pair of glial-like neuronal support cells in the tail, the phasmid sheath cells (PHsh). Amphid socket cells are labelled AMso, and phasmid socket cells are labelled PHso.

Figure 4. Ciliary and socket cell structures in the *bgnt-1* mutant are present and appear superficially wild-type except for a marginal increase in ADL cilium length. (a) Both amphid and phasmid cilia appear superficially wild-type. GFP-tagged CHE-2 (mammalian IFT80 orthologue) is used as a pan-cilia marker which localises to the basal bodies (bb) and axonemes. (b) ADL cilia correctly enter the amphid socket (AMso) cells in *bgnt-1* mutants. ADL cilia are labelled with P*srh-220*::IFT-20::tdTomato. IFT-20 (IFT20) localises to cilia basal bodies (bb) and axonemes. The *srh-220* promoter drives expression primarily in ADL neurons. The amphid socket (AMso) cells are labelled with cytoplasmic GFP driven by an amphid socket specific promoter, *grd-15* (Hunt-Newbury et al. 2007). **(c)** ADL cilia are significantly longer in *bgnt-1* mutants compared to wild-type ($p < 0.01$, Kruskal-Wallis test). **(d)** ADL cilia penetrate the sockets cells to an equivalent depth in wild-type and *bgnt-1* mutants ($p > 0.05$, Kruskal-Wallis

test). **(e)** ADL cilia/dendrites exhibit no guidance defects in *bgnt-1* mutants as assessed by the number of double-rod cilia observed in each amphid when ADL was driven by the primarily ADL specific *srh-220* promoter ($p > 0.05$, Fisher's exact test). Error bars represent 95% confidence intervals (Pearson Clopper method). **(f)** *bgnt-1* mutants do not exhibit a significant increase in the proportion of ADL neurons with dendritic blebbing ($p > 0.05$, Fisher's exact test). Error bars represent 95% confidence intervals (Pearson Clopper method).

Table 1. Summary of dye-fill phenotype classes observed

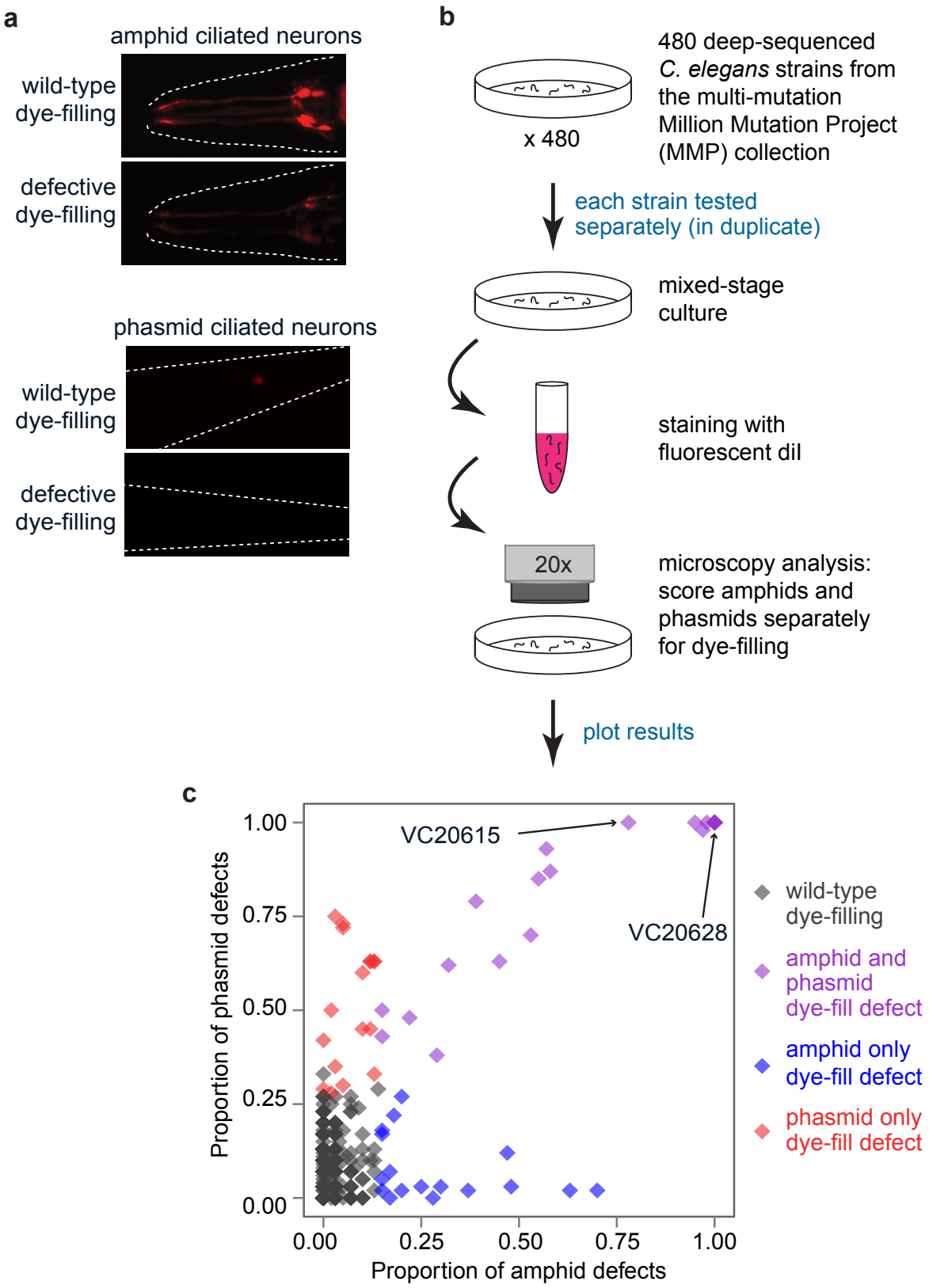
Phenotype summary	**Number of strains**
Amphid and phasmid dye-fill defect	11
Amphid and phasmid partial dye-fill defect	12
Phasmid only partial dye-fill defect	17
Amphid only partial dye-fill defect	17

Table 2. Genes with genome-wide significance for amphid ciliated neuron dye-filling phenotypes, ordered by p-value.

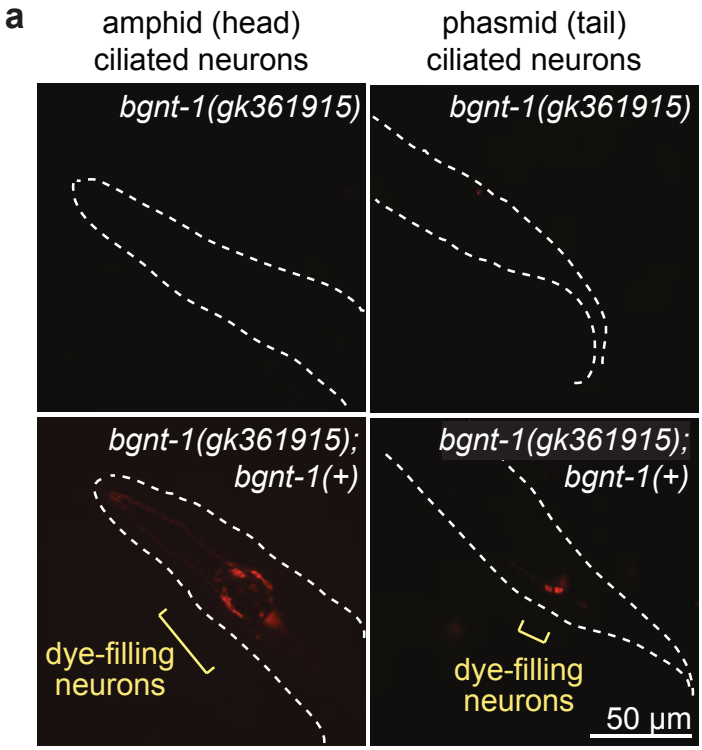
Gene	p-value	q-value	effect size (odds ratio)	# variants	Human homologue	function	ciliated sensory neuron evidence	*C. elegans* ciliated neuron expression
osm-1	3.81E-05	0.03	4.8	14	IFT172	Intraflagellar Transport complex B component	Perkins et al., 1986; Signor et al., 1999	yes
che-3	3.24E-04	0.11	1.9	21	DYNC2H1	Intraflagellar Transport dynein heavy chain	Perkins et al., 1986; Signor et al., 1999	yes
mdf-1	5.49E-04	0.14	12.1	8	MAD1	Mitotic spindle assembly checkpoint protein		unknown
bgnt-1	8.27E-04	0.14	7.1	7	B3GNT1	Glycoslytransferase		unknown
lars-1	8.86E-04	0.14	12.1	8	LARS	Leucyl-tRNA synthetase		unknown
cnt-1	7.01E-04	0.13	12.1	8	ACAP2	Arf-GAP	Jensen *et al*., Submitted	unknown
unc-80	2.67E-03	0.23	1.2	21	UNC80	Ion channel regulator		yes
F43D9.1	4.22E-03	0.26	8.8	7	PTCHD4	Patched domain-containing protein		yes
asp-1	4.36E-03	0.26	8.8	7	PGC	Progastricsin		unknown
H06I04.5	5.06E-03	0.26	8.8	7	none	Protein-tyrosine phosphatase-like		unknown
C05D12.2	5.13E-03	0.27	8.8	7	none	unknown		unknown
rbc-1	5.91E-03	0.27	4.3	19	DMXL1	DMX-like protein		unknown
itr-1	7.37E-03	0.28	5.3	13	ITPR1	Inositol 1,4,5-trisphosphate receptor		no
jip-1	7.75E-03	0.29	7.1	8	JIP	JNK interacting protein		unknown
gmps-1	7.85E-03	0.29	7.1	8	GMPS	GMP synthetase		unknown
fbxa-136	8.07E-03	0.29	6.1	15	none	F-box A protein		unknown
dcr-1	8.97E-03	0.29	6.9	11	DICER	Bidentate ribonuclease		unknown
jac-1	9.02E-03	0.29	6.9	11	CTND	Catenin delta		unknown

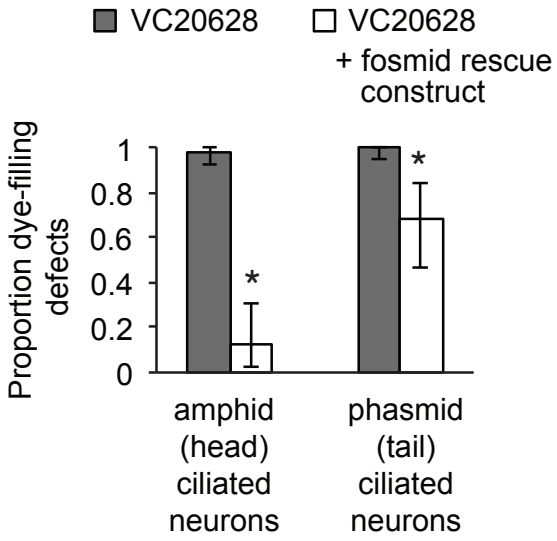
Table 3. Genes with genome-wide significance for phasmid ciliated neuron dye-filling phenotypes, ordered by p-value.

Gene	**p-value**	**q-value**	**effect size (odds ratio)**	**# variants**	**Human homologue**	**function**	**ciliated sensory neuron evidence**	***C. elegans* ciliated neuron expression**
osm-1	3.94E-05	0.04	5.24	14	IFT172	Intraflagellar Transport (IFT) complex B component	Perkins et al., 1986; Signor et al., 1999	yes
bgnt-1	1.21E-04	0.06	13.3	7	B3GNT1	Glycoslytransferase		unknown
che-3	3.71E-04	0.12	2.1	21	DYNC2H1	IFT dynein heavy chain	Perkins et al., 1986; Signor et al., 1999	yes
cnt-1	6.52E-04	0.16	13.2	8	ACAP2	Arf-GAP with coil coil, ankryin repeat and PH domains	Jensen et al., Submitted	unknown
R06F6.8	1.97E-03	0.25	8.8	10	RIC1	RAB6A GEF complex partner		unknown

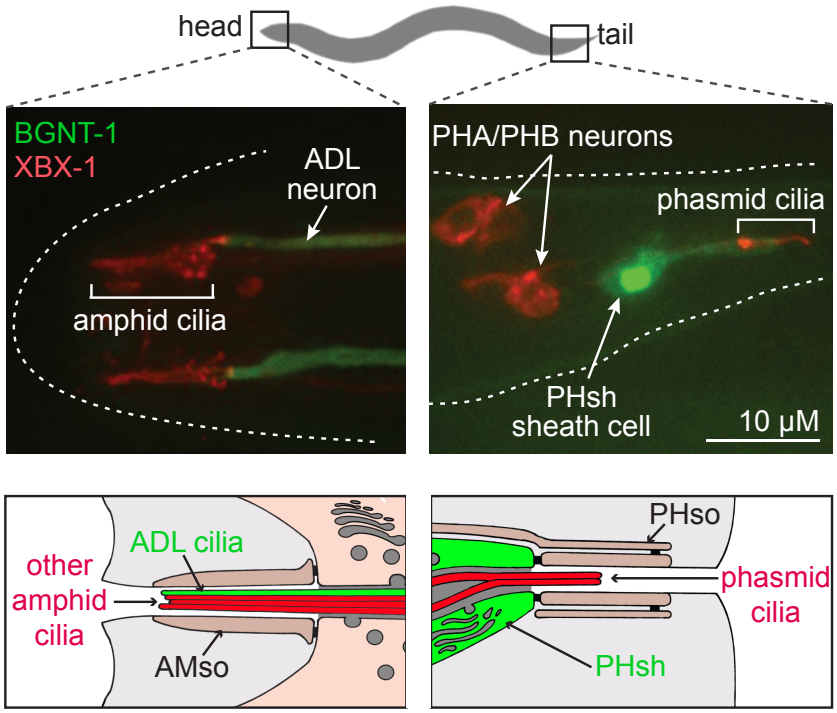


Timbers et al., Figure 1

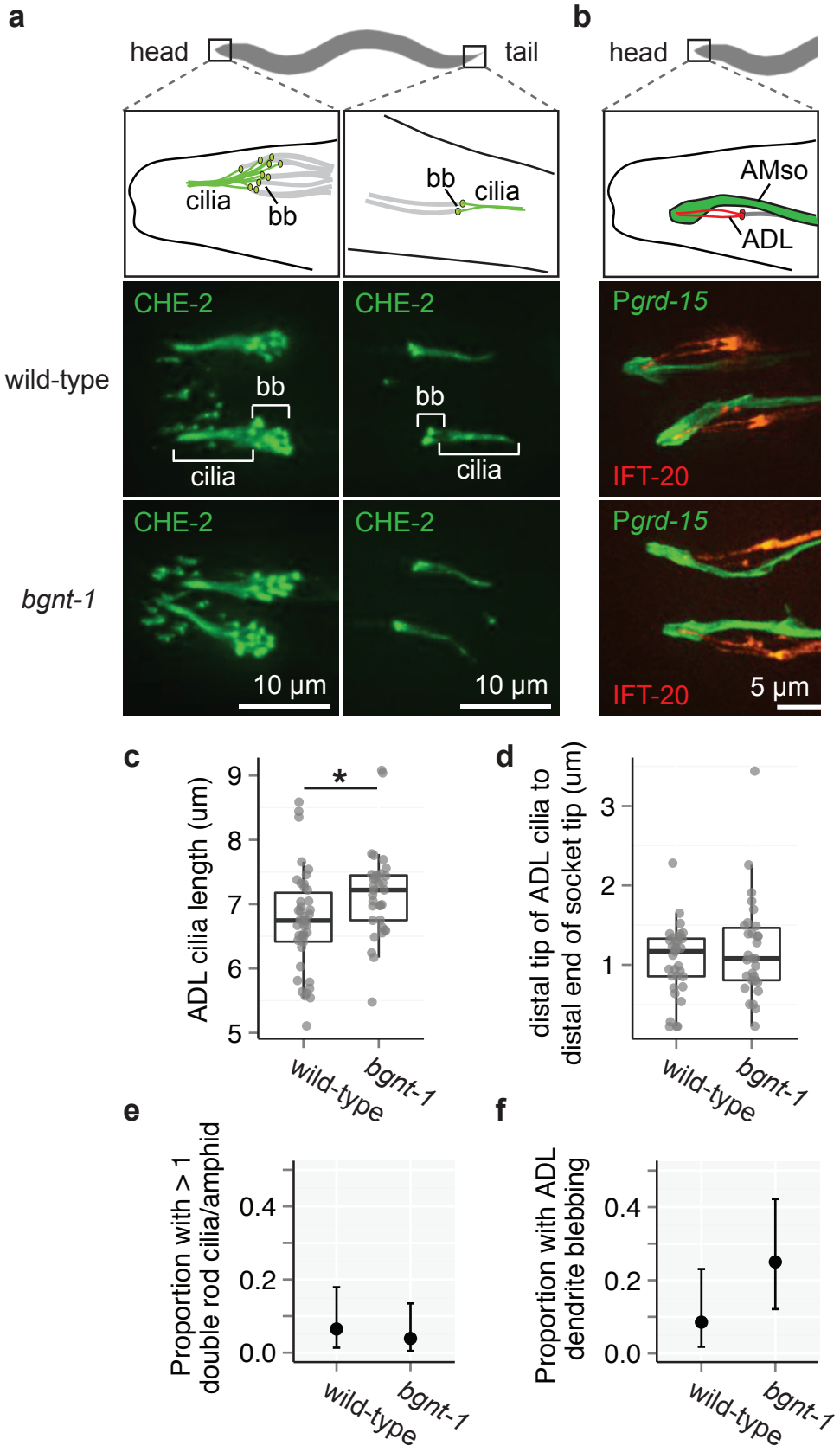




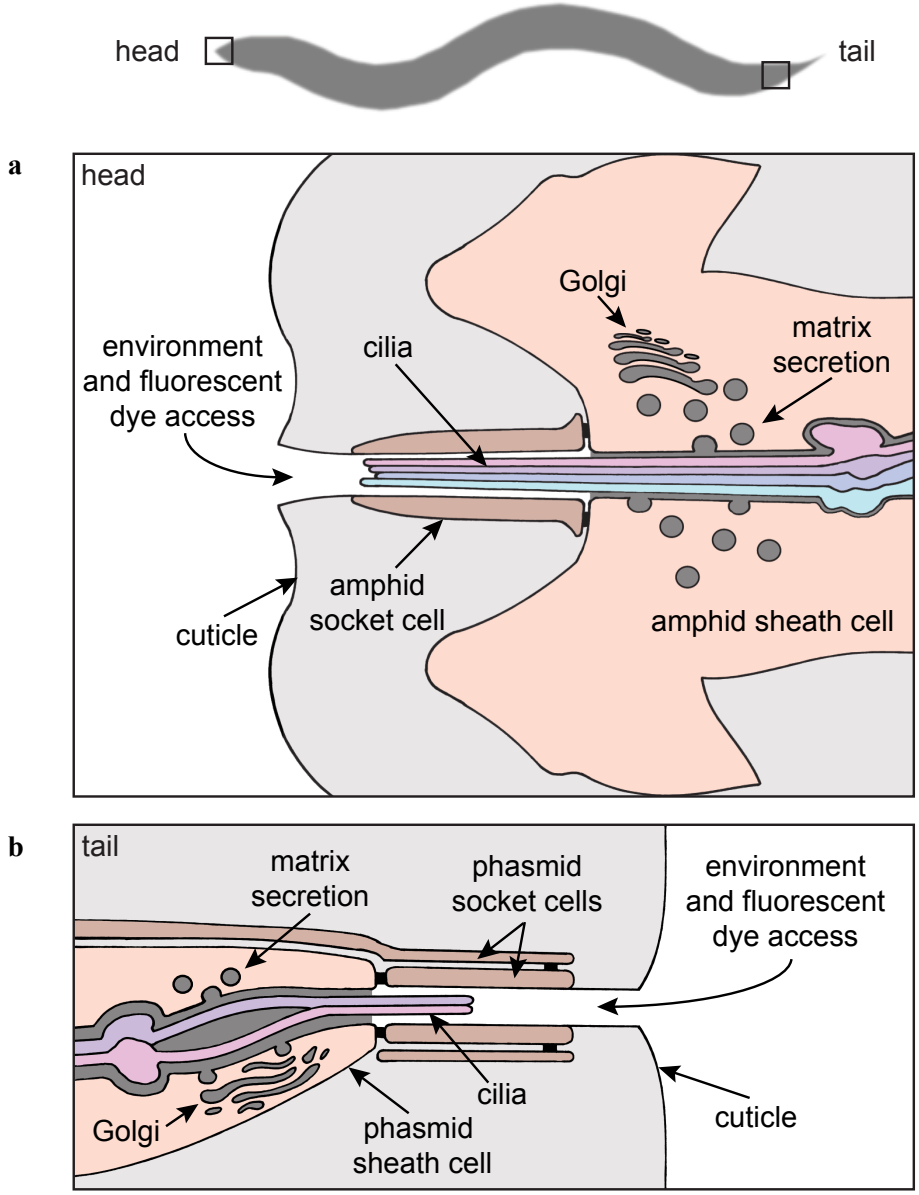
Timbers et al., Figure 2



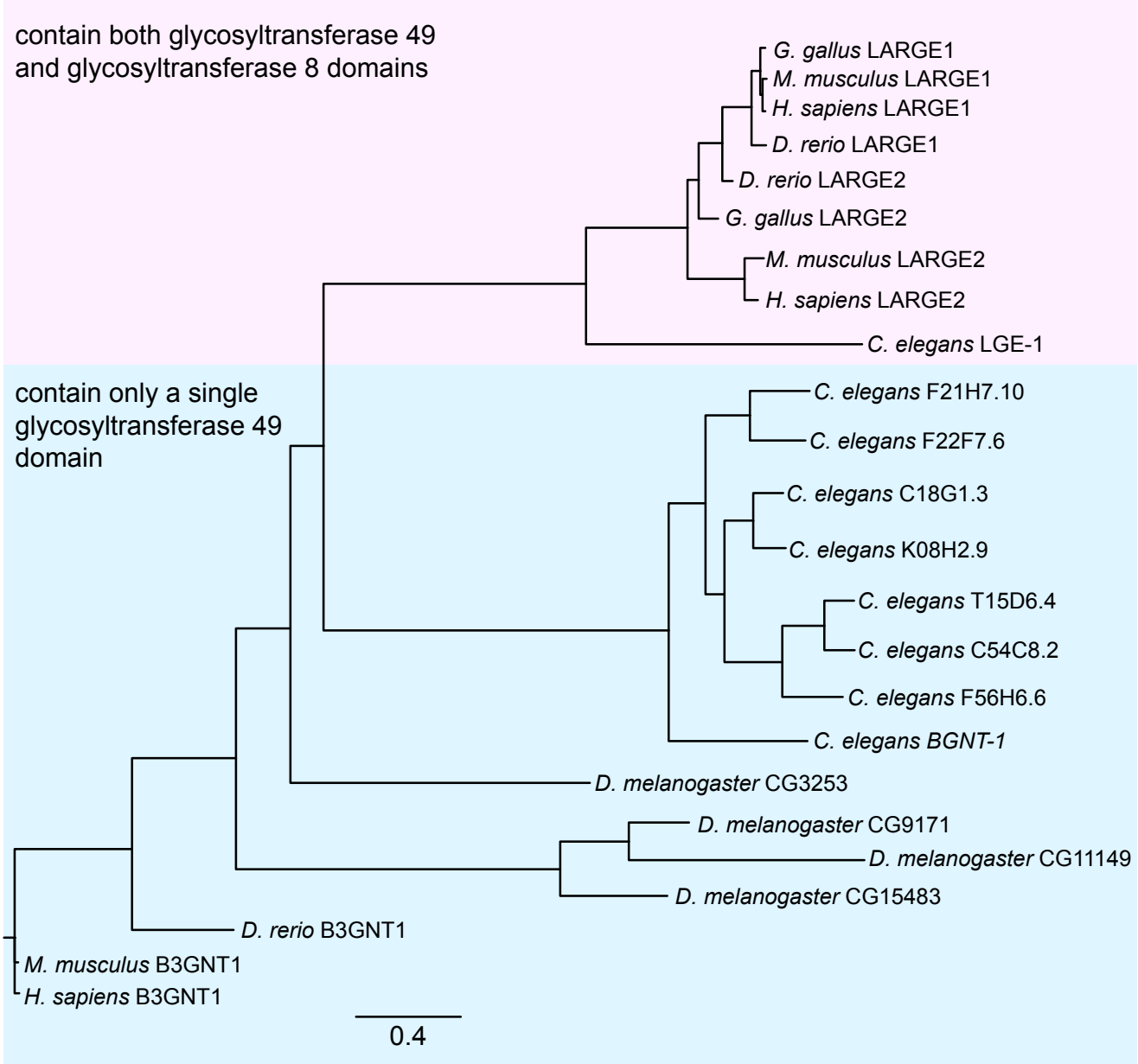
Timbers et al. Figure 3



Timbers et al. Figure 4



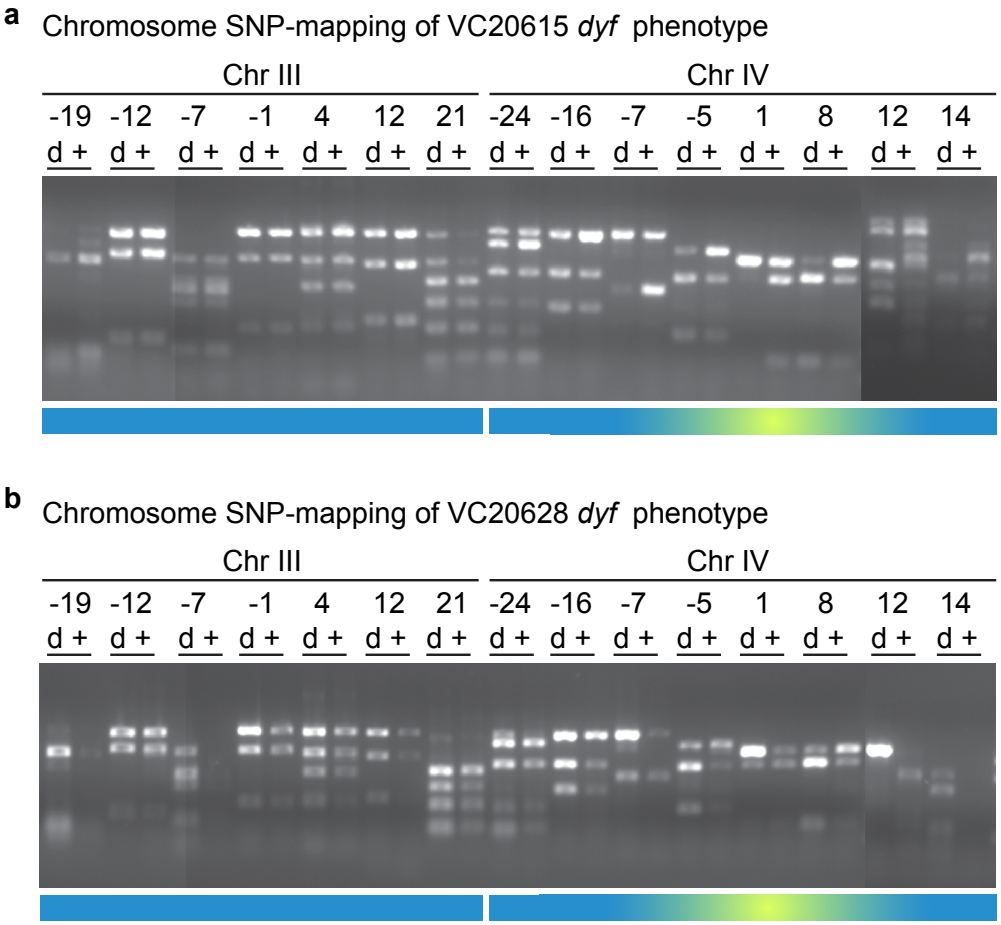
Supplementary Figure 1. Schematic of a longitudinal section through the wild-type amphid and phasmid sensillum. (a) The socket and sheath cells form the amphid channel through which the cilia are exposed to the environment. The lipophilic dye, DiI, is also presumed to access the cilia via this channel. Cuticle, continuous with the external surface of the worm, lines the socket channel. Matrix is secreted by the Golgi apparatus of the sheath cell, and fill the space that surrounds the cilia. (b) The phasmid sensillum is organised in the same manner as described for the amphid sensillum.



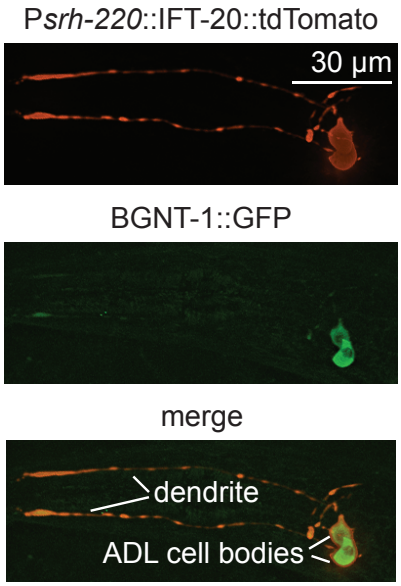
Supplementary Figure 2. Phylogenetic tree of glycosyltransferase 49 domain containing proteins. The glycosyltransferase 49 domain containing proteins appear to have radiated in *C. elegans.* BGNT-1 is the most basal member of this family, and most closely related to vertebrate B3GNT1. Protein sequences (obtained from: http://www.cazy.org/) were aligned using MUSCLE 3.7 (Edgar, 2004), and the phylogenetic tree was built using PhyML 3.0 aLRT (Guindon et al., 2010) and viewed using FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

```
C. elegans BGNT-1        MR-------GIARIIVHFLVLFQILCCVY--FATNMIRRKPGFRISDSS-GTSTKRNE--
D. melanogaster CG3253   MD--------------------------------------GFVVPNNS-ASLAFDNI--
H. sapiens B3GNT         MQMSYAIRCAFYQLLLAALMLVAMLQLLYLSLLSGLHGQEEQDQYFEFFPPSPRSVDQ--
D. rerio B3GNT           MH--FSKKCSVFKVVLSALLIVALLQLLYLSFLSKLHGKQQRYKYSELF-GSKKNANQGE
                         *                                            :    :    :

C. elegans BGNT-1        --TLGEYIRYESKKYGS-DFCVAYNV--TKATGNFRDDGLE-----P----ISLVLHATS
D. melanogaster CG3253   --NL---------SYGRWDNQLLYRIKDFALLGEQYVDSSE-------GSLVCLATQTSV
H. sapiens B3GNT         ---VKAQLRTALASGGVLDASGDYRV--YRGLLKTTMDPND----------VILATHASV
D. rerio B3GNT           KNPRREHLRYSLSTGGIFDGSGQYRV--YKNLIKSDFSTNQKPGADPRSHHLALATHTTI
                                    . *  *    *.:       :   .  :          : *. :::

C. elegans BGNT-1        HFMREIEGQCSSRNWNGPISISLFVDRTSSE--AVDYLHEVHRCSAKVNQKLSVHVVYRM
D. melanogaster CG3253   ERLNSLPQVA--ANWQGKMSVALFAAGPEEFVVLQYFVTYMRLCFANIRENATFHLLTPR
H. sapiens B3GNT         DNLLHLSGLL--ERWEGPLSVSVFAATKEEAQLATVLAYALSSHCPDMRARVAMHLVCPS
D. rerio B3GNT           NNLHHLESLL--ERWKNPISVAIF-ANGEDVKFATAIIYALSLFCPQVQALVDFHLVCHS
                         . :  :       .*:. :*:::*    ..          :    ..:.    .*::
                                                                                 P194S
C. elegans BGNT-1        S-----PFQKFCDPIL---IKRSNRKCTSFNSTI---RSRERSRVIPPFQ-----IYPIN
D. melanogaster CG3253   DFDKLPRVAALPLNMR------GKFDCQYPDRTL---KALLKFRSLKTLQWRQRNTYPQN
H. sapiens B3GNT         R-----YEAAVPDPREPGEFALLR-SCQEVFDKLARVAQPGINYALGTNV-----SYPNN
D. rerio B3GNT           G-----EMATFPDQDREHFVGLQEMGCPAVFAKLESHRDKYKNYAIGSNV-----SYPNN
                                   .            .  *     .:           : .        ** *
                                  G205E
C. elegans BGNT-1        VMRNVARKGALSYIHMTADVEMMFSDGFALKMKPLANKYINGKDKKLLVIRRFEVDNKAH
D. melanogaster CG3253   HMRNLARKGCQTKYVFLTDIDIVPSTNSVPQLNHFFRTA-NCTKSCAYVIPTFEIDVRAT
H. sapiens B3GNT         LLRNLAREG--ANYALVIDVDMVPSEGLWRGLREMLDQS-NQWGGTALVVPAFEIRRARR
D. rerio B3GNT           LLRNVARGGTDAAYILVIDIDMIPSANLHHQFVTMLMKR-EPAADEVLVLPAFEIRHIRK
                          :**:** *  :   :  *:::: * .    :  :     :       *:  **:

C. elegans BGNT-1        VPADNKELFLMIKAFRAFEFHHKYFPAGHTIESLWQWFRMSKNQTEAYAWKIDYKSSSWE
D. melanogaster CG3253   FPRSKNALVRLIRKGLARPFHEKVFIYNQYATNFSKWLSPNTNETEVS-----------V
H. sapiens B3GNT         MPMNKNELVQLYQVGEVRPFYYGLCTPCQAPTNYSRWVNLPEES--LL-----------R
D. rerio B3GNT           MPASKPELVQLYQVGEVRPFYDELCSRCQAPTNYSLWVNLASKSSGPL-----------E
                         .* .:  *. : :   .  *:       :   .   *.    :.

C. elegans BGNT-1        AQLILHRKDPYNPEYIPTR---IRDQQ---------SLVYELCRANYTFHLASHVFNVHR
D. melanogaster CG3253   SHVVTNFEFLYEPFYIAVDNAPAHDERFLGYGFTRNSQVYEMHIAGYQFYVLSPVFTGHW
H. sapiens B3GNT         PAYVVPWQDPWEPFYVAGGKVPTFDERFRQYGFNRISQACELHVAGFDFEVLNEGFLVHK
D. rerio B3GNT           VSYTINWVDPWEPFYIGARSVPLYDESFRQYGFNRISQACELHIAGYRFSVVSNAFLLHK
                                   ::* *:        *:          * . *:  *.: * : .  *  *

C. elegans BGNT-1        GVKTKETNLSSAVLTHQKRLRTRSYKRFMHYINTTYPDTFDQCGK------FVM
D. melanogaster CG3253   GLQRKQARPAWR--EQQNNANRRKFDVFKSEIFVRYKNDPRLLLKAKKNQILVK
H. sapiens B3GNT         GFKEALKFHPQK--EAENQHNKILYRQFKQELKAKYPNSPRRC-----------
D. rerio B3GNT           GFKVQGEFHSRK--DEENRKNRILFRSFKESLKAKYPTSPRRC-----------
```

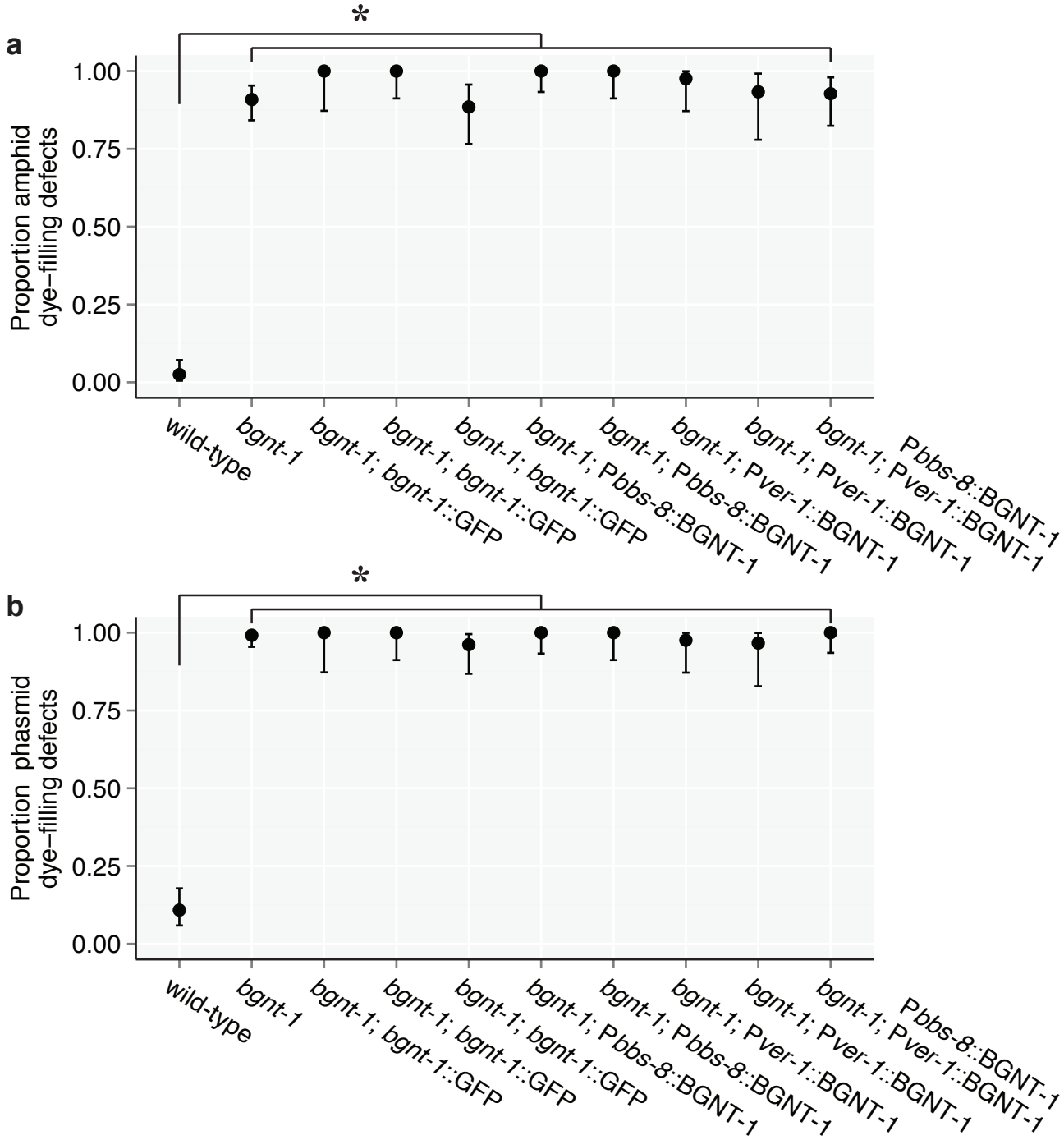
Supplementary Figure 3. Multiple sequence alignment of *C. elegans* BGNT-1 with homologous glycosyltransferase 49 domain containing proteins. In VC20615 the *bgnt-1(gk355974)* mutation results in P194S at a conserved P (highlighted in pink). In VC20628 *the bgnt-1(gk361915)* mutation results in G205E at a conserved E (highlighted in blue). Alignment was performed by L-INS-i MAFFT using BLOSUM62 as a scoring matrix and a gap onset penalty of 1.53.



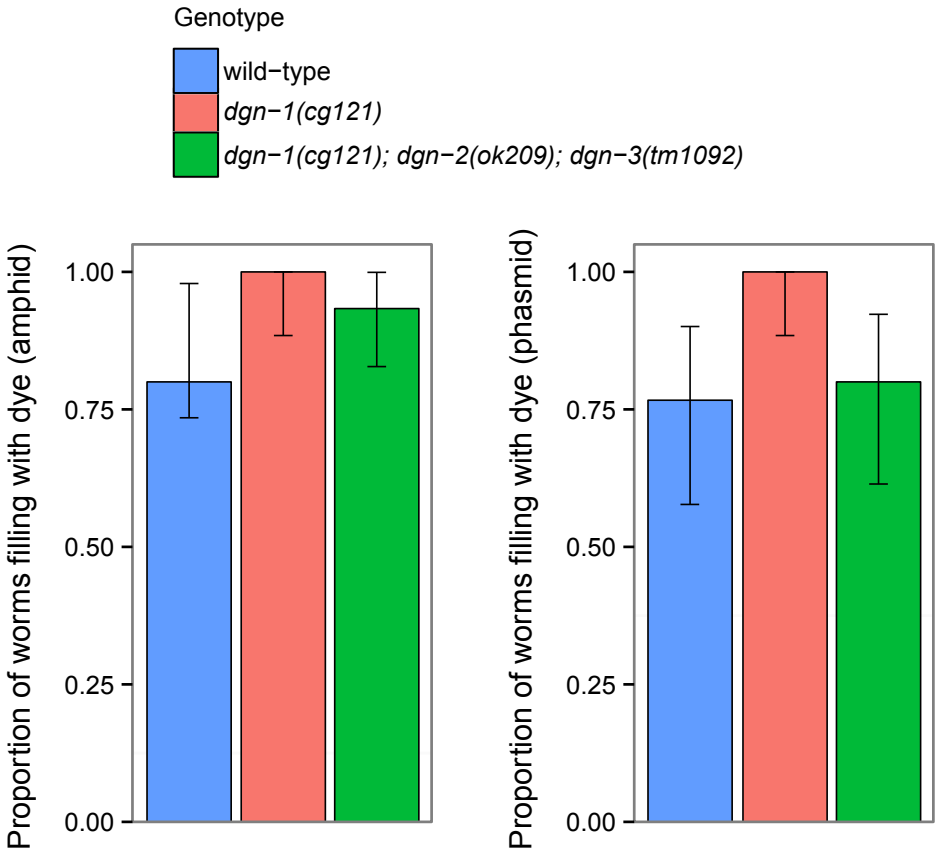
Supplementary Figure 4. VC20615 and VC20628 dye-fill defects map to the *bgnt-1* locus. (a) Chromosome SNP mapping of VC20615's dyf phenoype. Each pair of lanes shows results from the SNP at the indicated genetic map position, using DNA template from either worms exhibiting dyf (d) or wild-type (+) dye-filling. Linkage is visible as an increase in the proportion of Bristol N2 DNA in dyf lanes compared to the wild-type lanes, and is visible on ChrIV from -5 to 8. (b) chromosome mapping of VC20628's dyf phenoype. Similar to VC20615, linkage is visible as an increase in the proportion of Bristol N2 DNA in dyf lanes compared to the wild-type lanes, and is visible on ChrIV from -5 to 8. All PCR samples from SNP-mapping of each strain were run on the same large gel from which multiple images were captured to visualize all samples.



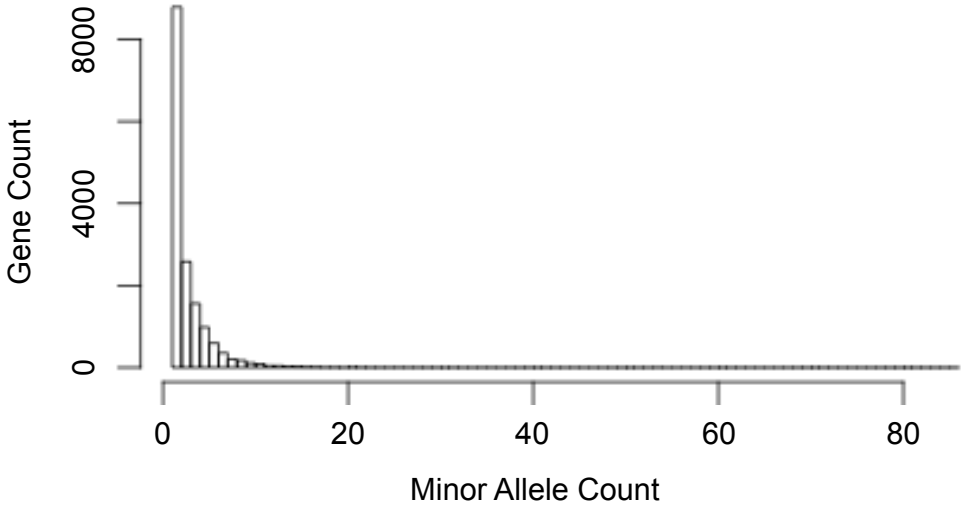
Supplementary Figure 5. Localisation of *C. elegans* BGNT-1. A translational C-terminal fusion of GFP with the endogenous *bgnt-1* promoter and coding region localises diffusely to a pair of ciliated sensory neurons in the head (ADL left and right). This is confirmed by colocalisation of the cell-specific ADL marker, P*srh-220*::IFT-20::tdTomato.



Supplementary Figure 6. Expressing *bgnt-1* cDNA under control of the endogenous *bgnt-1* promoter (3346 bp), the *bbs-8* pan-cilia promoter, or the *ver-1* sheath cell promoter as an extrachromosomal array does not rescue the dye-filling defects observed in *bgtn-1(gk361915)* mutants. Only wild-type worms filled with dye, all other strains exhibited significant dye-filling defects compared to wild-type ($p < 0.005$), but were not signficantly different from *bgnt-1* mutants ($p > 0.05$; Fisher's exact test followed by a Bonferroni correction). Error bars represent 95% confidence intervals (Clopper Pearson method).



Supplementary Figure 7. *C. elegans* Dystroglycan homologues, *dgn-1*, *dgn-2* and *dgn-3*, are not required for dye-filling of amphid or phasmid ciliated sensory neurons. dgn mutants did not exhibit significant dye-filling defects when compared to wild-type ($p > 0.05$, Fisher's exact test followed by a Bonferroni correction) Error bars represent 95% confidence intervals (Clopper Pearson method).



Supplementary Figure 8. Distribution of minor allele count for mutated genes from the 480 strains screened in this study from the Million Mutation project. The majority of genes have only a single mutation in the 480 strains screened.