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

Email addresses:
Tiffany A. Timbers - tiffany.timbers@gmail.com
Stephanie J. Garland - stephjgarland@gmail.com
Swetha Mohan - swethamohan@gmail.com
Stephane Flibotte - sflibotte@gmail.com
Mark Edgley - edgley@mail.ubc.ca
Quintin Muncaster - q.muncaster@gmail.com
Donald G. Moerman - moerman@zoology.ubc.ca
Michel R. Leroux - leroux@sfu.ca

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

Abstract (max 100 words)

Forward and reverse genetic screens represent powerful methods to uncover new components for any biological process, but suffer from gene cloning or throughput bottleneck issues, respectively. Here, we employ an innovative approach to gene discovery: we screened a *C. elegans* whole-genome sequenced multi-mutation library for our phenotype of interest, namely defective ciliated sensory neuron development, and performed the Sequence Kernel Association Test (SKAT) to rapidly identify genes exhibiting this phenotype. Our approach unveiled a novel gene, *bgnt-1.1*, that influences cilia length. Notably, the human orthologue *B3GNT1*/*B4GAT1* is associated with Walker-Warburg syndrome, a suspected ciliary disorder.

Keywords: SKAT, GWAS, genetic screen, sensory biology, Million Mutation Project, cilia, *bgnt-1.1*, B3GNT1, B4GAT1, *C. elegans*

Background

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 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; 1], 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 affecting nearly all *C. elegans* coding regions. On average, each strain contains ~ 400 non-synonymous mutations affecting protein coding sequences.

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; 2] allows us to effectively and efficiently predict novel genes important for our chosen biological process—the development and function of ciliated sensory neurons. Using this approach, we found that a previously uncharacterized gene, *bgnt-1.1*, plays an essential role in this process. Interestingly, *bgnt-1.1* is the orthologue of human B3GNT1/B4GAT1, a gene implicated in Walker-Warburg syndrome, a disorder with clinical ailments resembly a ciliary disease (ciliopathy).

Results

Primary (non-motile) cilia arise from a modified centriole (basal body) and act as 'cellular antennae' that transduce environmental cues to the cell [3]. They enable sensory physiology (such as olfaction/chemosensation, mechanosensation, vision) and are central to signalling pathways essential for metazoan development [4]. 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 [5, 6].

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**) [7, 8]. Many dye-filling (*dyf*) mutants known from genetic screens [8, 9] harbour mutations in genes influencing ciliated sensory neuron development and function, including ciliogenesis [10], cilia maintenance [11], axon guidance [9], dendrite anchoring/formation [12], cell fate [13], and neural support (glial) cells [14].

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 [2].

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 [15], 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; 16] and the combined multivariate and collapsing (CMC) method [17], 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 0.25. 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 19 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 6 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-1.1), as well as Arf- and Rab-related proteins (CNT-1 and RBC-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, and a Ral GTPase activating protein subunit (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 five genes associated with both amphid and phasmid dye-filling defects, namely *osm-1*, *che-3*, *cnt-1, rbc-1*, and *bgnt-1.1*, 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 [18]. 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 [18, 19]. 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, two of the other amphid/phasmid dye-filling gene hits, *cnt-1* and *rbc-1*, encode proteins that play roles 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 [3] is well established. *cnt-1* encodes the orthologue of human ACAP2, which interacts with both Rab35 [20] and Arf6 [21] to mediate crosstalk between these two proteins, at least in the context of PC12 cell neurite outgrowth, and potentially through endocytic recycling [22]. *rbc-1* is orthologous to human DMXL1 and DMXL2; the latter protein, also called Rabconnectin-3a (Rbcn-3a), regulates the endocytosis and exocytosis functions of Rab3a by way of its associations with WDR7/ Rabconnectin-3b (Rbcn-3b), RAB3GAP and RAB3GEF [23-25]. Notably, nonsense and splicing mutations in RABGAP1, the gene encoding the catalytic subunit of RAB3GAP, lead to Warburg Micro syndrome—a disorder characterised by eye and brain defects, as well as microgenitalia in males [26]. This disorder has been remarked by Baker and Beales [5] to have phenotypic overlap with 3 core features common to ciliopathies (including hypoplasia of the corpus callosum, mental retardation and polydactyly), indicating that it may derive from a possible ciliary pathology.

A third, putative novel dye-filling gene significantly associated with both amphid and phasmid dye-fill phenotypes is *bgnt-1.1* (**Table 2 & 3, Suppl. Table 3 & 4**). *bgnt-1.1* 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 [27]. In HeLa cells, its subcellular localisation is concentrated at the *trans*-Golgi [28]. B3gnt1 knockout mice exhibit axon guidance phenotypes [29, 30] and deficient behavioural responses to estrous females [31]. In humans, mutations in B3GNT1 are associated with a congenital muscular dystrophy with brain and eye anomalies, Walker-Warburg syndrome (WWS) [32, 33]. 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* [5]. Additionally, one patient is reported to exhibit dysplastic kidneys [32], 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-1.1 in dye-filling, and analyse its involvement role in ciliated sensory neuron development.

Of the eight MMP strains harbouring mutations in *bgnt-1.1*, three (VC20615, VC20628 and VC20326) exhibited severe dye-fill phenotypes (**Suppl. Table 1**). The C -> T missense mutation in *bgnt-1.1* in VC20615 corresponds to P194S alteration in the protein sequence, while VC20628 and VC20326 each harbour an identical G -> A missense mutation in *bgnt-1.1* 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-1.1* was responsible for the dye-filling phenotypes in *bgnt-1.1* mutants we rescued the dye-fill defects by expressing a fosmid containing a wild-type copy of *bgnt-1.1* in an extrachromosomal array (**Fig. 2a,b**). Another way to confirm that disruption of *bgnt-1.1* causes dye-fill defects would be to observe this phenotype in a strain harbouring a knock-out mutation in *bgnt-1.1*. Although there are 49 *bgnt-1.1* alleles available, a knock-out allele of *bgnt-1.1* does not yet exist. There are two insertion/deletion alleles available, *gk1221* and *tm4314*, but both fall within introns and likely do not affect protein function. Thus, we also tested the causality of *bgnt-1.1 via* a relatively efficient SNP mapping approach. We established that the dye-fill phenotypes from VC20615 and VC20628 strains mapped to the *bgnt-1.1* locus, on chromosome IV between -5 cM and 8 cM (**Suppl. Fig. 4**). Notably, in both VC20615 and VC20628 strains, *bgnt-1.1* was the only gene in this region harbouring a mutation which was common to both of these strains. Together, these findings indicate that the *bgnt-1.1* mutations in VC20615, VC20628 and VC20326 are causative for the observed dye-filling defects in these strains.

To shed light on how *bgnt-1.1* affects dye-filling, we expressed GFP-tagged BGNT-1.1, 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. 2c, Suppl. Fig. 5**). Promoters of ~ 1 kb (the length of the intergenic region between *bgnt-1.1* and the upstream gene), and 3346 bp (300 bp upstream of the upstream gene) both resulted in the same expression pattern of BGNT-1.1::GFP. This and several cell-specific rescue constructs was unable to rescue the dye-filling phenotype of *bgnt-1.1* mutants (**Suppl. Fig. 6**), indicating that the

expression pattern of this gene may be wider than we can visualize with a GFP reporter and/or that the function of BGNT-1.1 is sensitive to dosage (*C. elegans* rescues experiments generally result in some level of over-expression due to expressing transgenes *via* an extrachromosomal array).

Next, we queried whether *bgnt-1.1* mutants exhibit any gross ciliary morphology defects by expressing a ciliary marker in *bgnt-1.1* mutants, namely the GFP-tagged IFT-B subcomplex protein, CHE-2 (IFT80). This experiment indicated that although the cilia of *bgnt-1.1* mutants fail to fill with dye, their ciliary structures appear superficially wild-type (**Suppl. Fig. 7a**). Since modest cilia structure defects may be more difficult to observe using pan-cilia markers, due to overlapping ciliary signals, we also characterised the phenotype of cilia and dendrites in *bgnt-1.1* mutants within a single ciliated amphid cell, the ADL neuron. For this purpose, we used the primarily cell-specific ADL promoter, P*srh-220*, to drive expression of another cilia marker, IFT-20 (IFT20) tagged with tdTomato. 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-1.1* mutants (**Suppl. Fig. 7b**). We then sought to assay for potential phenotypes involving ADL cilia length (**Fig. 2d**); 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; **Suppl. Fig. 7c)**; ADL guidance (proportion of double rod cilia/amphid; **Suppl. Fig. 7d**); and finally, ADL dendrite blebbing

(structural alteration where dendrites take bead on a string appearance; **Suppl. Fig. 7e**). Our analyses revealed that ADL cilia in *bgnt-1.1* mutants are wild-type in most aspects except for a modest cilia length defect. Specifically, *bgnt-1.1* mutants were observed to have significantly longer cilia compared to wild-type worms (**Fig. 2d**; $p < 0.01$, Kruskal-Wallis test).

BGNT-1.1 therefore influences amphid and phasmid neuron development and cilium length, without overtly affecting the gross structure of neurons or cilium formation. How disrupting BGNT-1.1 abrogates dye filling remains uncertain. One possibility is that the glycosyltransferase regulates the association of cilia with the sheath and socket glial-like cells which envelop them [7] (**Suppl. Fig. 1**). This defect will 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 [7]. Which substrate(s) the β1,3-N-acetylglucosaminyltransferase, BGNT-1.1 (B3GNT1), glycosylates, and how this influences sensory neuron/glial cell development and function, remains to be determined in a future, detailed study of the gene.

Discussion

Here we demonstrate that rare-variant 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-1.1* were significantly associated with dye-filling defects, suggesting that a significant proportion of the remaining 17

genes significantly associated with this phenotype likely represent genes important for ciliary/ sensory neuron development and/or function.

We confirmed that *bgnt-1.1,* a gene SKAT identified as being associated with the dye-filling phenotypes but not previously implicated in cilia or amphid-sensillum function, is a bona-fide dye-filling gene. We observed that: *(1)* two missense mutations in *bgnt-1.1* result in severe dye-fill defects in 3 MMP strains; *(2)* a fosmid containing full-length wild-type *bgnt-1.1* rescues the dye-filling phenotype in *bgnt-1.1* mutants; *(3)* the dye-filling phenotypes in the MMP strains with mutations in *bgnt-1.1* map to the *bgnt-1.1* locus; *(4)* BGTN-1 is expressed in cells critical for dye-filling; and finally; and *(5)* mutations in *bgnt-1.1* result in a small but statistically significant ciliary length defect. Together, these data strongly indicate that BGNT-1.1 functions in dye-filling. Further experiments to determine the mechanism of how mutations in *bgnt-1.1* lead to this phenotype will be included in a future study focused on BGNT-1.1.

In humans, mutations in *B3GNT1* cause Walker-Warburg syndrome [WWS; 32, 33]. 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. 8**), indicating that BGNT-1.1 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 [5]. Our findings that *C. elegans bgnt-1.1* 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.

We performed SKAT analyses via two methods, 1) while applying functional weights to the variants, and 2) while weighting all variants equally. SKAT analysis of the 480 strains with and without weights resulted in a similar list of genes not previously known to be associated with the dye-filling phenotype, although interestingly, SKAT without weights did not result in the known dye-filling genes, *osm-1* and *che-3*, being significantly associated with dye-filling defects (**Supp. Table 5 & 6**). We hypothesise the reason for this discrepancy is that without weights, the effect of gene size (odds ratio) on dye-filling for these genes is small. This is a consequence of the many non-deleterious MMP missense mutation alleles for *osm-1* and *che-3* 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-1.1*), 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 [34] or Polyphen [35] 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) t*esting 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 [1], as well as partially-penetrant or other difficult-to-score phenotypes. In the case of *bgnt-1.1* we performed genetic rescue experiments and SNP mapping to support our the SKAT findings, which indicates that *bgnt-1.1* 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-1.1*, 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. 36, 37, 38] 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 [39]. 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. 9**), 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.

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 [40]. 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 [1]. The Million Mutation Project strains were isolated and their genomes' sequenced by Thompson *et al*. [1]. 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-1.1.

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*. [41]. The following SNPs used by Davis *et al*. [41] were omitted from our analysis because the whole genome sequence data from Thompson *et al.* [1] 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-1.1(gk361915)*, 25 ng/µl of fosmid WRM065bB05 containing *bgnt-1.1* was injected into *bgnt-1.1* mutants along with 80 ng/µl of pRF4

rol-6(su1006dm) as a co-injection marker. *bgnt-1.1(gk361915)*; Ex[CHE-2::GFP; pRF4] was created by crossing *bgnt-1.1(gk361915)* with wild-type worms expressing Ex[CHE-2::GFP; pRF4]. The translational Psrh-220::IFT-20::tdTomato fusion was generated as described in [11], 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-1.1(gk361915)* was also introduced to this line via genetic crossing to create MX2236 *bgnt-1.1(gk361915); dpy-5(e907)*; Ex[P*srh-220*::IFT-20::tdTomato; pRF4]; sIs12964[P*grd-15*::GFP; pCeh361]. The translational BGNT-1.1::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-1.1* gene, removing the *bgnt-1.1* 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-1.1 cell specific rescue construct was created by PCR fusion to stitch 841 bp of the *bbs-8* promoter to the bgnt-1.1 cDNA and 920 bp of the *unc-54* 3' UTR. The P*ver-1*::BGNT-1.1 cell-specific rescue construct was created by using PCR fusion to stitch 1982 bp of the *ver-1* promoter to the bgnt-1.1 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-1.1 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-1.1. 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) [2] 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 [1], 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

Competing interests

The authors declare that they have no competing interests.

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-1.1*. 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.

Acknowledgements

T.A.T. acknowledges a Banting Postdoctoral Fellowship from the Canadian Institutes of Health Research (CIHR). M.R.L. acknowledges funding from 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).

Description of additional data files

The following additional data are available with the online version of this paper.

Additional data file 1: Supplementary Figure 1. Schematic of a longitudinal section through the wild-type amphid and phasmid sensillum.

Additional data file 2: Supplementary Figure 2. Phylogenetic tree of glycosyltransferase 49 domain containing proteins.

Additional data file 3: Supplementary Figure 3. Multiple sequence alignment of *C. elegans* BGNT-1.1 with homologous glycosyltransferase 49 domain containing proteins.

Additional data file 4: Supplementary Figure 4. VC20615 and VC20628 dye-fill defects map to the *bgnt-1.1* locus.

Additional data file 5: Supplementary Figure 5. Localisation of *C. elegans* BGNT-1.1 to the ADL neurons.

Additional data file 6: Supplementary Figure 6. Expressing *bgnt-1.1* cDNA under control of the endogenous *bgnt-1.1* 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.

Additional data file 7: Supplementary Figure 7. Ciliary and socket cell structures in the *bgnt-1.1* mutant are present and appear superficially wild-type.

Additional data file 8: Supplementary Figure 8. *C. elegans* Dystroglycan homologues, *dgn-1*, *dgn-2* and *dgn-3*, are not required for dye-filling of amphid or phasmid ciliated sensory neurons.

Additional data file 9: Supplementary Figure 9. Distribution of minor allele count for mutated genes from the 480 strains screened in this study from the Million Mutation project.

Additional data file 10: Supplementary Table 1. List of the dye-filling phenotype of each MMP strain.

Additional data file 11: Supplementary Table 2. List of the strains which exhibit a dye-fill defect and the known dye-filling genes which are mutated in these strains.

Additional data file 12: Supplementary Table 3. List of the results from the SKAT analysis of the amphid dye-filling phenotype using functional weights.

Additional data file 13: Supplementary Table 4. List of the results from the SKAT analysis of the phasmid dye-filling phenotype using functional weights.

Additional data file 14: Supplementary Table 5. List of the results from the SKAT analysis of amphid dye-filling phenotype where all variants were weighted equally.

Additional data file 15: Supplementary Table 6. List of the results from the SKAT analysis of phasmid dye-filling phenotype where all variants were weighted equally.

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 [1]. 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-1.1,* a gene identified by SKAT to be associated with both amphid and phasmid dye-filling defects.

Figure 2. Confirmation of *bgnt-1.1* as a novel dye-filling gene.

(a) Wild-type *bgnt-1.1* rescues dye-filling defects in *bgnt-1.1 (gk361915)* mutants. Transformation of VC20628 *bgnt-1.1 (gk361915)* with a fosmid containing wild-type *bgnt-1.1*

(WRM065bB05) completely rescues the amphid, and partially rescues the phasmid dye-filling defects of VC20628 *bgnt-1.1 (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). **(c)** *C. elegans* BGNT-1.1 localises to ADL ciliated sensory neurons and phasmid sheath cell. A translational C-terminal fusion of GFP with the endogenous *bgnt-1.1* 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). Cilia and ciliated neurons are labelled with the well characterized P*osm-5*::XBX-1::tdTomato marker. Amphid socket cells are abbreviated to AMso, and phasmid socket cells are abbreviated to PHso. **(d)** ADL cilia are significantly longer in *bgnt-1.1* mutants compared to wild-type ($p < 0.01$, Kruskal-Wallis test). 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.

References

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

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

3. Sung CH, Leroux MR: The roles of evolutionarily conserved functional modules in cilia-related trafficking. Nat Cell Biol 2013, 15:1387-1397.

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

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

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

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

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

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

10. Inglis PN, Ou G, Leroux MR, Scholey JM: The sensory cilia of Caenorhabditis elegans. WormBook 2007:1-22.

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

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

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

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

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

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

17. Li B, Leal SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008, 83:311-321.

18. Signor D, Wedaman KP, Orozco JT, Dwyer ND, Bargmann CI, Rose LS, Scholey JM: 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. J Cell Biol 1999, 147:519-530.

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

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

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

22. Kobayashi H, Fukuda M: Rab35 regulates Arf6 activity through centaurin-beta2 (ACAP2) during neurite outgrowth. J Cell Sci 2012, 125:2235-2243.

23. Nagano F, Kawabe H, Nakanishi H, Shinohara M, Deguchi-Tawarada M, Takeuchi M, Sasaki T, Takai Y: Rabconnectin-3, a novel protein that binds both GDP/GTP exchange protein and GTPase-activating protein for Rab3 small G protein family. J Biol Chem 2002, 277:9629-9632.

24. Kawabe H, Sakisaka T, Yasumi M, Shingai T, Izumi G, Nagano F, Deguchi-Tawarada M, Takeuchi M, Nakanishi H, Takai Y: A novel rabconnectin-3-binding protein that directly binds a GDP/GTP exchange protein for Rab3A small G protein implicated in Ca(2+)-dependent exocytosis of neurotransmitter. Genes Cells 2003, 8:537-546.

25. Li KW, Chen N, Klemmer P, Koopmans F, Karupothula R, Smit AB: Identifying true protein complex constituents in interaction proteomics: the example of the DMXL2 protein complex. Proteomics 2012, 12:2428-2432.

26. Aligianis IA, Johnson CA, Gissen P, Chen D, Hampshire D, Hoffmann K, Maina EN, Morgan NV, Tee L, Morton J, et al: Mutations of the catalytic subunit of RAB3GAP cause Warburg Micro syndrome. Nat Genet 2005, 37:221-223.

27. Zhou D, Berger EG, Hennet T: Molecular cloning of a human UDP-galactose:GlcNAcbeta1,3GalNAc beta1, 3 galactosyltransferase gene encoding an O-linked core3-elongation enzyme. Eur J Biochem 1999, 263:571-576.

28. Lee PL, Kohler JJ, Pfeffer SR: 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 2009, 19:655-664.

29. Henion TR, Raitcheva D, Grosholz R, Biellmann F, Skarnes WC, Hennet T, Schwarting GA: Beta1,3-N-acetylglucosaminyltransferase 1 glycosylation is required for axon pathfinding by olfactory sensory neurons. J Neurosci 2005, 25:1894-1903.

30. Bless E, Raitcheva D, Henion TR, Tobet S, Schwarting GA: Lactosamine modulates the rate of migration of GnRH neurons during mouse development. Eur J Neurosci 2006, 24:654-660.

31. Biellmann F, Henion TR, Burki K, Hennet T: Impaired sexual behavior in male mice deficient for the beta1-3 N-acetylglucosaminyltransferase-I gene. Mol Reprod Dev 2008, 75:699-706.

32. 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: Missense mutations in beta-1,3-N-acetylglucosaminyltransferase 1 (B3GNT1) cause Walker-Warburg syndrome. Hum Mol Genet 2013, 22:1746-1754.

33. Shaheen R, Faqeih E, Ansari S, Alkuraya FS: A truncating mutation in B3GNT1 causes severe Walker-Warburg syndrome. Neurogenetics 2013.

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

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

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

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

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

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

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

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

Table 1. Summary of dye-fill phenotype classes observed

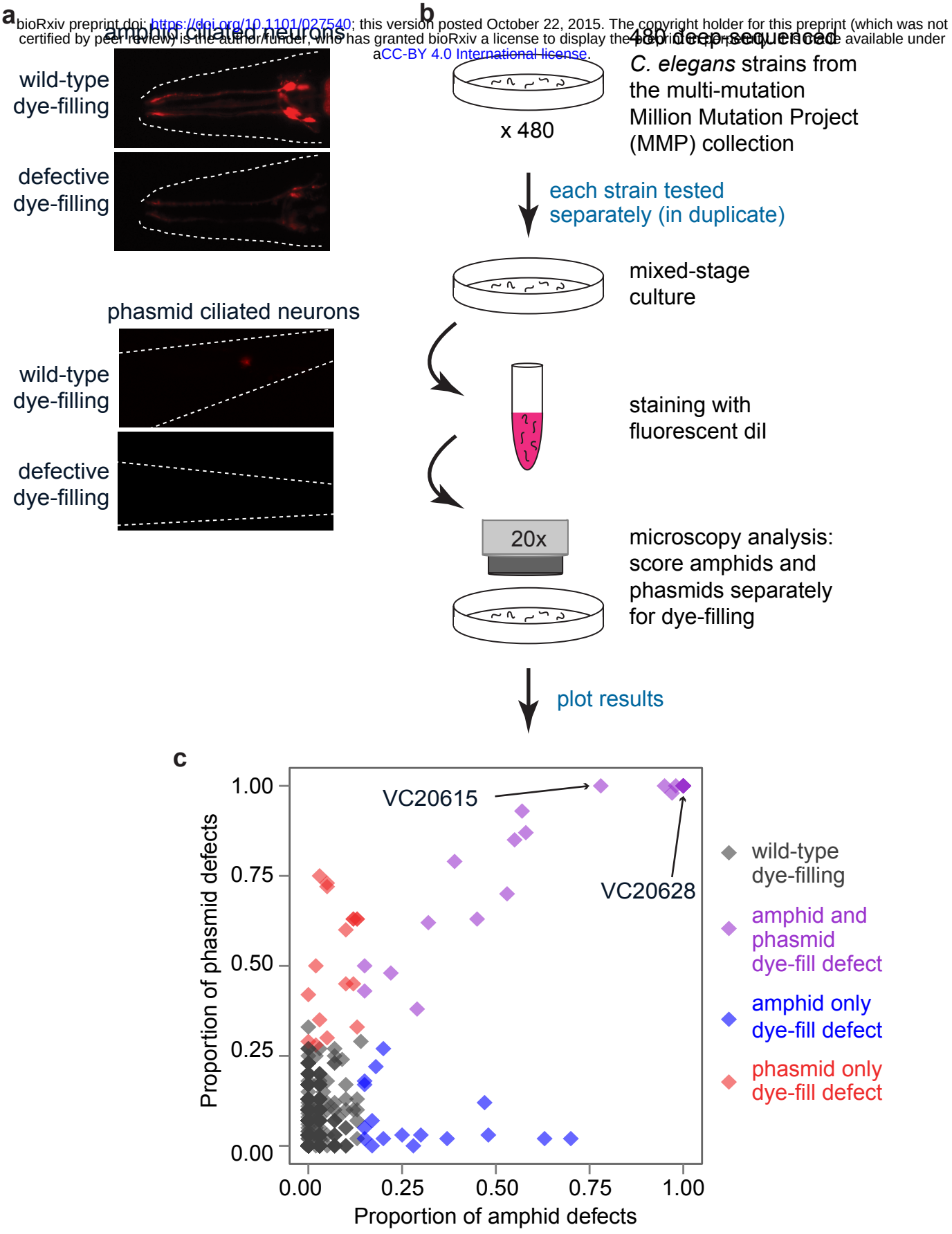
Phenotype summary	**Number of strains**
Amphid and phasmid dye-fill defect	**11**
Amphid and phasmid partial dye-fill defect	**12**
Phasmid only partial dye-fill defect	**17**
Amphid only partial dye-fill defect	**17**

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

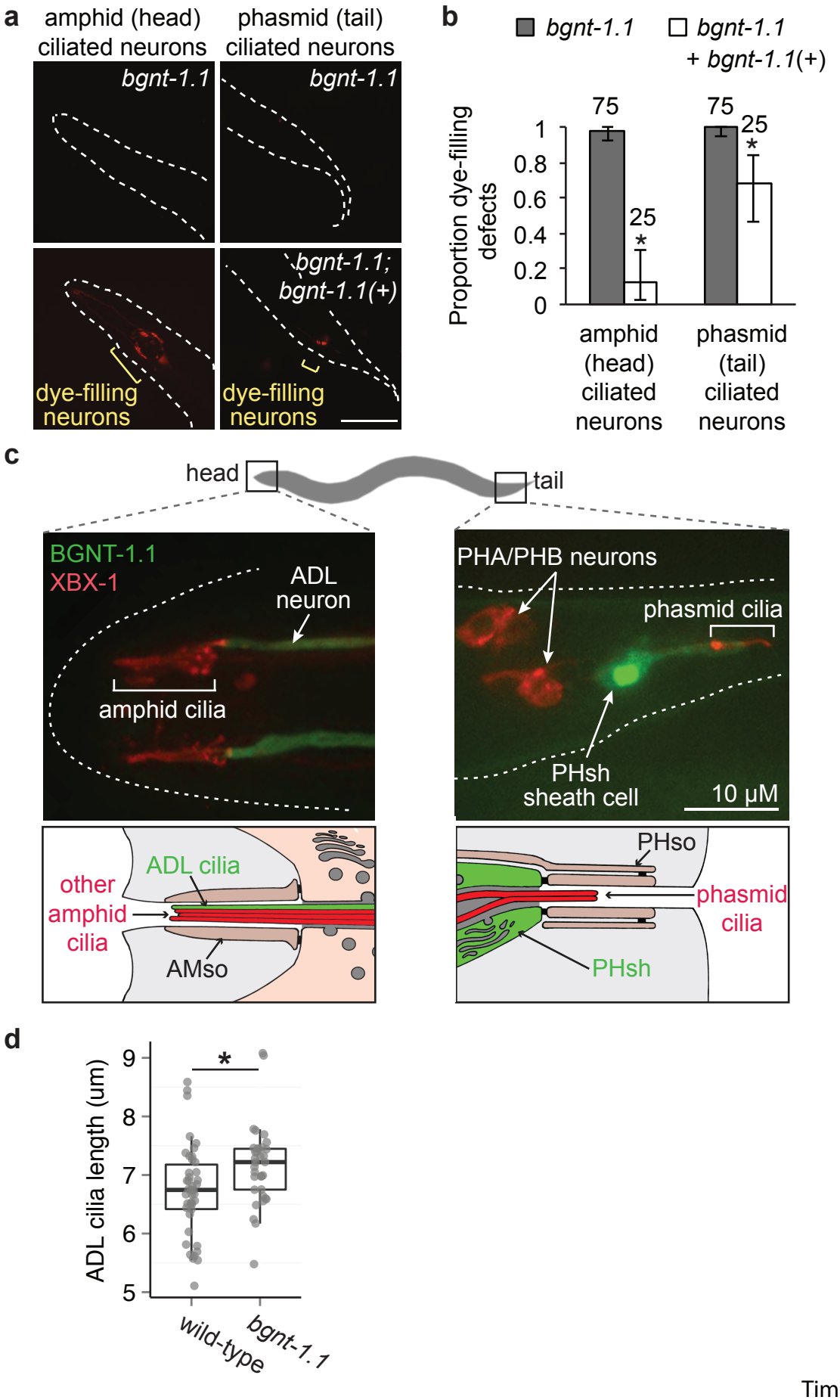
Gene	Sequence	p-value	q-value	effect size (odds ratio)	# variants	Human homologue	function	ciliated sensory neuron evidence	*C. elegans* ciliated neuron expression
osm-1	T27B1.1	3.1E-05	0.03	4.8	14	IFT172	Intraflagellar Transport complex B component	Perkins et al., 1986; Signor et al., 1999	yes
che-3	F18C12.1	3.8E-04	0.11	1.9	21	DYNC2H1	Intraflagellar Transport dynein heavy chain	Perkins et al., 1986; Signor et al., 1999	yes
cnt-1	Y17G7B.15	6.1E-04	0.12	12.1	8	ACAP2	Arf-GAP	Jensen *et al*., Submitted	unknown
lars-1	R74.1	6.1E-04	0.12	12.1	8	LARS	Leucyl-tRNA synthetase		unknown
mdf-1	C50F4.11	5.0E-04	0.12	12.1	8	MAD1	Mitotic spindle assembly checkpoint protein		unknown
bgnt-1.1	F01D4.9	1.0E-03	0.16	7.1	7	B3GNT1	Glycoslytransferase		unknown
unc-80	F25C8.3	2.6E-03	0.23	1.2	21	UNC80	Ion channel regulator		yes
F43D9.1	F43D9.1	4.5E-03	0.26	8.8	7	PTCHD4	Patched domain-containing protein		yes
C05D12.2	C05D12.2	5.1E-03	0.27	8.8	7	none	unknown		unknown
H06I04.5	H06I04.5	5.1E-03	0.27	8.8	7	none	Protein-tyrosine phosphatase-like		unknown
asp-1	Y39B6A.20	5.4E-03	0.27	8.8	7	PGC	Progastricsin		unknown
rbc-1	F54E4.1	6.8E-03	0.28	4.3	19	DMXL1	DMX-like protein		unknown
jip-1	F56D12.4	7.5E-03	0.28	7.1	8	JIP	JNK interacting protein		unknown
fbxa-136	C18D4.2	7.6E-03	0.28	6.1	15	none	F-box A protein		unknown
gmps-1	M106.4	7.7E-03	0.28	7.1	8	GMPS	GMP synthetase		unknown
dcr-1	K12H4.8	8.4E-03	0.29	6.9	11	DICER	Bidentate ribonuclease		unknown
itr-1	F33D4.2	8.6E-03	0.29	5.3	13	ITPR1	Inositol 1,4,5-trisphosphate receptor		no
jac-1	Y105C5B.21	9.0E-03	0.29	6.9	11	CTND	Catenin delta		unknown
hgap-2	D2085.5	1.1E-02	0.29	4.4	11	RALGAPB	Ral GTPase activating protein subunit		no

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

Gene	Sequence	p-value	q-value	effect size (odds ratio)	# variants	Human homologue	function	ciliated sensory neuron evidence	*C. elegans* ciliated neuron expression
osm-1	T27B1.1	4.5E-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	F01D4.9	1.7E-04	0.08	13.3	7	B3GNT1	Glycoslytransferase		unknown
che-3	F18C12.1	3.0E-04	0.10	2.1	21	DYNC2H1	IFT dynein heavy chain	Perkins et al., 1986; Signor et al., 1999	yes
cnt-1	Y17G7B.15	7.9E-04	0.17	13.2	8	ACAP2	Arf-GAP with coil coil, ankryin repeat and PH domains	Jensen et al., Submitted	unknown
R06F6.8	R06F6.8	1.5E-03	0.23	8.8	10	RIC1	RAB6A GEF complex partner		unknown
rbc-1	F54E4.1	4.0E-03	0.29	4.8	19	DMXL1	DMX-like protein		unknown



Timbers et al., Figure 1



Timbers et al., Figure 2