Lineage-specific control of convergent cell identity by a Forkhead repressor

Karolina Mizeracka[1,2], Julia M. Rogers[3,4], Shai Shaham[5], Martha L. Bulyk[3,4,6], and Maxwell G. Heiman[1,2,*]

[1] Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA

[2] Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA

[3] Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[4] Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA

[5] The Rockefeller University, New York, NY

[6] Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

*heiman@genetics.med.harvard.edu

ABSTRACT

A central goal in developmental biology is to decipher the molecular events that govern cell fate specification in each developmental lineage. Here, we show that the *C. elegans* Forkhead transcription factor UNC-130 specifies two glial types that arise from one lineage, but does not affect equivalent glia that are produced in different anatomical regions from other lineages. We show that glial defects correlate with UNC-130:DNA binding, and that UNC-130 acts as a transcriptional repressor via two independent domains. UNC-130 can be functionally replaced by its human homolog, the neural crest lineage determinant FoxD3, and other neural crest factors (UNC-86/Brn3 and RNT-1/Runx) act in the same pathway. We propose that, in contrast to "terminal selectors", UNC-130 acts as a "lineage selector" to enable molecularly distinct progenitor cells to generate regionally equivalent cell types. This novel mechanism may underlie the recent observation of convergent lineages as a prevalent feature of vertebrate development.

INTRODUCTION

Embryonic development is the story of how a single, totipotent cell produces cell lineages that ultimately give rise to every cell type in the body. A framework for understanding this process has relied on two assumptions: first, that cell lineages diverge like tree branches to produce increasingly dissimilar cell states and, second, that each cell state can be reached only by a unique series of branches. One known exception to this paradigm has been the neural crest, a specialized lineage only present in vertebrates, that generates highly diverse cell types, including progeny that adopt the same identity as cells that are produced from completely unrelated lineages – a process termed "convergent differentiation" (Keyte and Hutson, 2012). Recently, single-cell RNA profiling has revealed more examples that challenge our basic assumptions (Briggs et al., 2018; Cao et al., 2019; Farrell et al., 2018; Pijuan-Sala et al., 2019; Raj et al., 2018; Wagner et al., 2018). Tracking of cell lineage histories in conjunction with single cell profiling has shown that, in fact, unrelated progenitors can frequently give rise to molecularly similar cell types, suggesting that convergent differentiation is surprisingly common during development (Chan et al., 2019; Liu et al., 2019; McKenna and Gagnon, 2019; Wagner et al., 2018). These observations imply that, contrary to previous assumptions, identical cell states can be reached by more than one specification trajectory. Because single-cell profiling studies are inherently descriptive, a major remaining challenge is to identify lineage-specific regulatory mechanisms that govern convergent differentiation.

C. elegans provides a powerful system for interrogating the relationships between embryonic origins and final cell fates, as it has an invariant lineage that has been fully

mapped (Sulston et al., 1983), and reporters are available for many unique cell types. Similar to vertebrate embryogenesis, a number of distinct sublineages in *C. elegans* produce equivalent cell types in different anatomical regions (Packer et al., 2019; Sulston et al., 1983). We focused on glia as a model to examine how regional cell type specification is controlled. Most *C. elegans* glia are located in sense organs in the head – the inner labial (IL), outer labial (OL), cephalic (CEP), and amphid (AM) sensilla – each of which contains two glial cell types, called the sheath and socket. Most sense organs are four or six-fold radially symmetric, such that each glial type is present in different regions. For example, there are three pairs of IL socket glial cells – dorsal, lateral, and ventral (ILsoD, ILsoL/R, ILsoV, respectively) – all of which express the same reporter genes and appear as a uniform transcriptional cluster in single-cell profiling experiments (Packer et al., 2019). Importantly, in many cases, equivalent glial types in different anatomical positions arise from unrelated lineages.

Here, we elucidate a lineage-specific role for the conserved Forkhead transcription factor UNC-130. During *C. elegans* development, UNC-130 is expressed in a multipotent sublineage that produces sensory neurons and glia, and was previously shown to be required for the specification of sensory neurons (Sarafi-Reinach and Sengupta, 2000), suggesting it establishes a specific cell identity. In contrast, we find that UNC-130 is required broadly throughout this sublineage, including for the specification of glial cell types, but it is not required for the production of equivalent glia that arise from unrelated lineages. Thus, unlike transcription factors that act in a cell-type-specific manner, UNC-130 controls the specification of diverse cell types by acting at the level of a cell lineage. The vertebrate homolog of UNC-130, FoxD3, also acts in a

lineage-specific manner and is required for the specification of neural crest-derived cell types (Kos et al., 2001; Lister et al., 2006; Stewart et al., 2006). Intriguingly, we find that UNC-130 and FoxD3 share molecular features, including their preferred binding sites, and FoxD3 can functionally replace UNC-130. Our results highlight a novel example demonstrating that cells that appear as a single mature "type" actually arise developmentally by convergent paths from distinct lineages which require different regulatory factors. Lineage-specific regulatory factors like UNC-130/FoxD3 may have been overlooked in systems where it is difficult to determine the relationship between lineage and cell fate, but may represent an evolutionarily ancient mechanism for cell fate specification.

RESULTS

UNC-130 acts in a sublineage to specify glial subtypes

To identify genes controlling glial fate, we performed a chemical mutagenesis screen and isolated a mutant strain with abnormal sense organ morphology (see Methods). Genetic mapping and sequencing revealed a causal mutation in the *unc-130* gene, which encodes a conserved Forkhead transcription factor. We obtained the reference allele for this gene, the deletion strain *ev505*, and characterized sense organ perturbations in detail (Nash et al., 2000). We find that the majority of *unc-130* mutants (95%) lost expression of an ILso reporter (*grl-18*pro) specifically in dorsal (ILsoD), but not lateral (ILsoL/R) or ventral (ILsoV) glial cells (Figure 1A, B, G; see Methods). Specification of the six associated IL neurons is unaffected; lack of specific reporters precluded examination of IL sheath glia.

We wanted to understand why mutation of *unc-130* affects the specification of the ILsoD pair of glia, but not the equivalent lateral or ventral pairs. We noted that the lineages that give rise to dorsal, or to lateral and ventral, ILso glia diverge at the 4-cell stage, when the AB blastomere divides into the progenitor cells ABa and ABp (Figure 1H) (Sulston et al., 1983). Previous studies showed that *unc-130* expression commences midway through embryogenesis in an ABp-derived lineage that gives rise to dorsal – but not lateral or ventral – ILso glia, as well as the amphid sheath (AMsh) glia, sensory neurons, and hypodermis (Figure 1H) (Murray et al., 2012). Interestingly, other cell types in this lineage are also mis-specified in *unc-130* mutants: the sensory neuron ASG adopts the fate of its sister neuron, AWA; and a cell specified to undergo programmed cell death ectopically acquires ASI neuronal fate (Sarafi-Reinach and Sengupta, 2000).

We therefore examined the remaining glial type in this sublineage, the bilaterally symmetric AMsh glia. Interestingly, we find that AMsh specification is also affected, with 22% of *unc-130* mutants failing to express an AMsh specific marker, *F16F9.3*, in one or occasionally both cells (Figure 1C, D). Loss of *F16F9.3* expression is highly correlated with loss of two other AMsh glia-specific reporters, *T02B11.3*pro:GFP and *F53F4.13*pro:GFP (Bacaj et al., 2008) (Figure S1). Importantly, there were no defects in the specification of phasmid sheath glia (PHsh), which are the functional equivalent of the AMsh glia in the tail, but are produced by an unrelated sublineage (Figure 1E, F, G). Other glia were also unaffected in *unc-130* mutants (Figure 1G, S1).

Taken together, our results suggest that UNC-130 promotes glial identity in a specific lineage, rather than regulating terminal specification of equivalent glial cells that arise from divergent lineages. Intriguingly, FoxD3, the UNC-130 vertebrate homolog,

also acts in a particular lineage (neural crest) to promote the specification of multiple cell types (peripheral neurons, glia, melanocytes, and others) (Lukoseviciute et al., 2018; Stewart et al., 2006; Teng et al., 2008).

Glial cells are mis-specified in *unc-130* mutants

Surprisingly, upon further examination of the amphid sense organ, we found that in addition to the usual pair of *grl*-2-expressing amphid socket glia (AMso), 42% of *unc-130* mutants have one or two extra *grl-2*+ cells (Figure 2A, S2). All ectopic *grl-2*+ cells also express the AMso-specific markers *itr-1* and *lin-48* (Figure 2B, S2). Furthermore, in wild-type males, each AMso glial cell undergoes a stem-cell-like division that maintains an AMso and generates a male-specific MCM neuron (Sammut et al., 2015). Unlike wild-type males, which have two MCMs, 20% of *unc-130* mutant males have three cells (Figure 2C, S2). Thus, the ectopic AMso cells in *unc-130* mutants express multiple AMso markers and exhibit similar progenitor properties as wild-type AMso cells.

We reasoned that the source of the ectopic AMso glia could be duplication of the cell lineage that generates AMso, or alternatively, these glia could arise directly by mis-specification of ILsoD. To distinguish between these possibilities, we examined male-specific CEM neurons, which are the sister cell of AMso. If the ectopic AMso arises by a lineage duplication, then we would expect to see additional CEM neurons as well; however, surprisingly, we found no difference in CEM number between wild-type and *unc-130* mutants (Figure S2). Furthermore, while AMso glia cell bodies are normally positioned laterally, we noted that the ectopic AMso glia cell bodies are always found dorsally, where ILsoD cell bodies are normally situated (100%, n=20) (Figure 2D-I).

Together, these observations suggest that ILsoD cells are likely mis-specified in *unc-130* mutants to the AMso glial fate. Intriguingly, the AMso fate may be a "default" state for socket glia, consistent with a hypothesis proposed previously to explain transformations of sensory neurons in *unc-130* mutants (Sarafi-Reinach and Sengupta, 2000).

UNC-130 DNA binding strength correlates with specification defects

We examined ILsoD and AMsh glial defects in existing *unc-130* mutant alleles (Nash et al., 2000; Sarafi-Reinach and Sengupta, 2000) and found that they could be binned into two categories (Figure 3A, B). In "strong" alleles, which include a point mutation in tryptophan 201 (W201G), an early stop mutant (R62stop), and a deletion allele, >90% of ILsoD glial cells lose reporter expression (average of 0.1-0.3 cells marked per animal) and exhibit extra AMso glia. In "weak" alleles, which include point mutations in arginine 218 or arginine 219 (R218C, R218H, R219K), 50% of ILsoD cells lose reporter expression (average of 0.9-1.2 cells marked per animal), and extra AMso glia are almost never observed. AMsh glia defects are, surprisingly, equally penetrant across all alleles (Figure 3B), suggesting a differential requirement for UNC-130 activity throughout this sublineage.

Although the W201 and R218-R219 amino acids lie within a short stretch of the primary protein sequence, mutations in these sites exhibit strikingly different phenotypic strength (Figure 3A, S3). As the relevant amino acids in UNC-130 are conserved in the DNA-binding domain (DBD) of its vertebrate homolog FoxD3, we mapped the position of these residues onto the structure of the FoxD3 DBD interacting with DNA (Jin et al., 1999) (Figure 3C, S3). The FoxD3 DBD contains four α-helices, one of which inserts

into the major groove of the DNA, two β-sheets, and two wing domains that stabilize DNA interaction (Jin et al., 1999). W201 is located in one of the β-sheets, whereas the R218 and R219 residues are in a wing domain on the opposite side of the helix bundle (Figure 3C).

Next, to understand the phenotypic binning of *unc-130* alleles, we examined the effects of the strong W201G and the weak R218C mutations on DNA binding *in vitro* (Figure 3C, S3). We first identified wild-type UNC-130-DBD DNA binding sites by incubating recombinant UNC-130-DBD protein with universal protein binding microarrays (PBMs) containing all possible 10-mer double-stranded DNA sequences (Berger et al., 2006). Previous PBM analysis of different metazoan Forkhead transcription factors uncovered three consensus binding sequences: a primary, canonical motif RYAAAYA (FkhP), a related secondary motif AHAACA (FkhS), or an alternate motif GACGC (FHL) (Nakagawa et al., 2013). We find that UNC-130-DBD preferentially binds [A/G][T/C]AAACA and AA[T/C]AACA sequences, variants of the primary and secondary Forkhead binding motifs, respectively, but not the alternate FHL motif (Figure 3D). To assess how UNC-130 binding specificity is affected by the W201G and R218C point mutations, we applied mutant recombinant proteins separately onto PBMs. We found that the R218C protein still bound DNA sequences recognized by wild-type UNC-130, but with lower PBM enrichment (E) scores, indicating lower affinity binding (Figure 3E, S3). The W201G mutant protein did not bind sequences recognized by wild-type protein, nor did it bind preferentially to any other motifs, suggesting severely impaired DNA binding (Figure 3E). Thus, moderately impaired DNA binding

appears to promote weak ILso defects, whereas severely impaired DNA binding results in strong ILso defects and appearance of extra AMso glia.

UNC-130 functions as a repressor to promote glial specification

The UNC-130 homolog FoxD3 acts as a repressor in several developmental contexts, in some cases recruiting the Groucho repressor complex through a conserved engrailed homology (eh1) domain (Ono et al., 2014; Yaklichkin et al., 2007a). Paradoxically, FoxD3 also functions as a pioneer factor in embryonic stem cells and during early neural crest specification (Krishnakumar et al., 2016; Lukoseviciute et al., 2018). To determine whether UNC-130 acts as a repressor or an activator in *C. elegans*, we expressed constructs encoding truncated versions of UNC-130 protein under the *unc-130* promoter in an *unc-130* mutant strain and assessed rescue of glia specification defects (Figure 4A). Expression of full-length UNC-130 fully rescues ILsoD and AMsh glia defects (Figure 4A). Expression of the UNC-130 DBD alone is not sufficient for rescue (Figure 4A). Addition of a transcriptional activation domain, VP64, to the DBD does not improve rescue, suggesting that endogenous UNC-130 is unlikely to act solely as a transcriptional activator. Expression of a protein containing the UNC-130 amino-terminus and DBD partially rescues ILsoD defects and fully rescues AMsh glia loss (Figure 4A). A protein lacking a candidate eh1 interaction motif in the UNC-130 amino-terminus fails to rescue either glial defect (Yaklichkin et al., 2007b) (Figure 4A, S3). Thus, like FoxD3, UNC-130 interaction with the Groucho repressive complex is necessary for UNC-130 function.

Expression of the DBD with the carboxy-terminus also rescues all glial specification defects (Figure 4A). Because the carboxy-terminus does not contain known transcriptional regulatory domains, we tested its activity in luciferase assays in cultured mammalian cells. We found that N-term:GAL4DBD:C-term and GAL4DBD:C-term have reduced reporter activity compared to GAL4 DBD alone (Figure 4B), suggesting the carboxy-terminus harbors repressive activity. Together, these studies provide evidence consistent with the idea that the UNC-130 carboxy- and amino- termini independently promote transcriptional repression *in vivo* to regulate the specification of multiple different types of glia arising from the same sublineage.

Requirements for fate specification in *C. elegans* are conserved in human FOXD3

The primary sequence and role of FoxD3 in specifying the neural crest lineage are highly conserved across vertebrates (Dottori et al., 2001; Kos et al., 2001; Lister et al., 2006; Sasai et al., 2001; Stewart et al., 2006; Teng et al., 2008). In contrast, *C. elegans* lacks a neural crest, and UNC-130 and human FOXD3 proteins share little sequence similarity outside the DBD and eh1 motif (Figure 5A, S3). Nonetheless, we found that expression of an *unc-130* promoter::*FOXD3* cDNA transgene almost completely rescues all glial specification defects in *unc-130* mutants (Figure 5A), suggesting that UNC-130 and FoxD3 functionality is conserved despite divergence at the primary sequence level.

To expand on this idea, we examined whether homologs of other neural crest determinants, Brn3 and Runx, are also involved in *C. elegans* glial specification. In vertebrates, Brn3 and Runx act downstream of FoxD3 and play important roles in specifying sensory neuron fates in the neural crest (Chen et al., 2006; Dottori et al., 2001;

Dykes et al., 2010; Eng et al., 2004; Kramer et al., 2006; Levanon et al., 2002; Lukoseviciute et al., 2018; Marmigère et al., 2006). Remarkably, although mutants in *C. elegans* UNC-86/Brn3 or RNT-1/Runx do not affect ILsoD, AMso, or AMsh specification in an otherwise wild-type background (Figure 5B, S4), loss of UNC-86 dramatically suppresses ectopic AMso glia generation in *unc-130* mutants (Figure 5B). Similarly, RNT-1 loss suppresses AMsh glia defects in *unc-130* mutants (Figure 5B).

In summary, while the neural crest is a uniquely vertebrate invention, the similarities between FoxD3 and UNC-130 extend from their roles in lineage specification to their molecular mechanisms of action – including their preferred DNA binding sites; their roles as transcriptional repressors, likely via an interaction with the Groucho repressive complex; and their genetic interactions with UNC-86/Brn3 and RNT-1/Runx. Our results suggest that, prior to the evolution of the neural crest, there already existed a FoxD3 precursor that acted in a lineage that undergoes both divergent and convergent differentiation.

DISCUSSION

Cell fate establishment is often viewed through the lens of "terminal selectors" – transcription factors expressed in post-mitotic cells that act as master regulators to establish and maintain a particular cell fate (Hobert and Kratsios, 2019). For example, the fate of all ILso glia is likely determined by the same terminal selectors. Intriguingly, although UNC-130 also acts to specify glia, its mode of activity is inconsistent with the definition of a terminal selector. First, UNC-130 is expressed in progenitor cells, and is not detected in post-mitotic, differentiating cells (Murray et al., 2012; Sarafi-Reinach and

Sengupta, 2000). Second, it does not specify a particular cell type, rather it acts in progenitor cells of a specific lineage to promote specification of diverse cells (glia, neurons, and a programmed cell death). This is clearly exemplified in this study, as UNC-130 is required by dorsal, but not lateral or ventral ILso glia. Likewise, it is required for AMsh glia specification but not the highly similar PHsh glia in the tail. Thus, we propose that UNC-130 serves as a "lineage selector," a transcription factor that acts in progenitor cells to promote specification of multiple cell types arising from a single lineage (see schematic in Figure 5C). Potentially, during development, branches of lineage specification pathways comprise different combinations of lineage selectors and terminal selectors, thereby creating unique divergent and convergent cell fate differentiation trajectories.

Because UNC-130 is required broadly for specification of the cells produced by this sublineage, it is unlikely to impart specific cell type information. Rather it may function to transition progenitor cells to a more restricted state and/or repress alternative fates. The lineages that give rise to the other IL socket glial pairs likely also utilize distinct but functionally similar factors to become differentiated. This may also be the case for other convergently-derived glial cells in *C. elegans*. For example, previous studies examining the specification of CEPsh glia showed that loss of *mls-2/*Nkx affects mainly ventral, but not dorsal, CEPsh glia (Yoshimura et al., 2008). Similar to the ILso glia, dorsal and ventral CEPsh glia are derived from lineages that diverge early in development, and it is possible that the observed phenotypes are also due to lineage-specific specification programs. These findings provide clear examples of how the same cell type can be produced by different molecular trajectories.

Addressing the question of how cell types are specified is particularly daunting in the nervous system, which is an intricate structure comprised of many diverse cell types arising from a myriad of lineages. Single-cell RNA profiling of the mature brain has shown that the glial classes of astrocytes and microglia, which had long been thought to consist of molecularly homogeneous cells, actually exhibit striking region-specific molecular heterogeneity (Hammond et al., 2019; John Lin et al., 2017; Marques et al., 2016; Masuda et al., 2019; Morel et al., 2017; Spitzer et al., 2019; Zeisel et al., 2015, 2018). In contrast, the molecular signatures of oligodendrocytes are highly similar (Marques et al., 2018; Zeisel et al., 2018). Because astrocytes and oligodendrocytes are thought to share a common progenitor, these observations suggest that regionally distinct progenitor cells may undergo convergent differentiation in the mammalian nervous system as well. Elucidation of these pathways will require single-cell profiling methods to be combined with careful lineage tracing and functional perturbation of regulatory factors *in vivo*, to ultimately achieve the level of resolution that is available in the far simpler nervous system of *C. elegans*.

ACKNOWLEDGMENTS

We thank Laura Moriarty for assistance with the genetic screen that isolated *ns313*; Oliver Hobert for whole genome sequencing; Piali Sengupta for *unc-130* cDNA and the *oy10* mutant strain; Joseph Culotti for *unc-130* mutant strains; Raphael Bruckner for assistance with cell culture; and Constance Cepko, Christopher Walsh, and the members of the Heiman laboratory for comments and advice on the manuscript. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

FUNDING

This project was supported by a William Randolph Hearst fellowship to K.M.; Bioinformatics and Integrative Genomics training grant T32HG002295 from NHGRI and National Science Foundation Graduate Research fellowship to J.M.R.; R01HG003985 and R01HG010501 from NIH/NHGRI to M.L.B; R35NS105094 from NIH/NINDS to S.S.; and a William F. Milton Award to M.G.H.

AUTHOR CONTRIBUTIONS

Conceptualization: K.M. and M.G.H.; Experimental design: K.M. and M.G.H.; PBM experiment and analysis: J.M.R; Writing of manuscript: K.M. and M.G.H. with input from all authors; Supervision and funding: S.S., M.L.B., and M.G.H.

COMPETING INTERESTS

M.L.B. is a co-inventor on patents on PBM technology. All other authors declare no other competing interests.

MATERIALS AND METHODS

Strains

Strains were constructed in the N2 background and cultured under standard conditions (Brenner, 1974). Transgenic strains were generated with standard techniques (Mello and Fire, 1995) with injection of 100 ng/µL of DNA (5-50µL per plasmid). Strains, transgenes, and plasmids are listed in Supp. Tables 1-3 respectively.

Isolation and mapping of *unc-130* alleles

We isolated an allele of *unc-130, ns313,* from a genetic screen for sense organ abnormalities. Animals of genotype *oyIs44* V were mutagenized using 70 mM ethyl methanesulfonate (EMS, Sigma) at 20°C for 4 hours. Nonclonal F2 progeny were examined on an Axioplan 2 fluorescence microscope (Zeiss) with 633/1.4 NA objective (Zeiss) and dual-band filter set (Chroma, Set 51019) and animals with sense organ defects were recovered. A mutant strain, *ns313*, exhibiting short amphid dendrites (14% penetrance) and amphid sheath migration defects (55% penetrance) was isolated.
With standard linkage mapping and SNP analysis (Wicks et al., 2001), *ns313* was mapped to an interval between -6 cM and 15 cM on LG II. *ns313* animals were crossed to the Hawaiian strain CB4856 and F2 progeny with the mutant phenotype were transferred to individual plates. All F3 recombinants were pooled and subjected to genomic DNA extraction and whole-genome sequencing for one-step mapping (Doitsidou et al., 2010). Analysis with CloudMap (Minevich et al., 2012) identified a linked region on LG II including a point mutation in *unc-130* (GAACTAT[T>G]GGGCGTGGA) (W201G).

Characterization of glial phenotypes

To score regional defects in IL sockets, we generated strains that co-expressed the IL socket marker (*grl-18*pro:mApple, *hmnIs47*) with a marker for URX (*flp-8*pro:GFP, *ynIs48*), a dorsally located neuron whose dendrites fasciculate with the processes of the dorsal, but not lateral or ventral, IL sockets as a landmark. Glial specification defects were scored visually on either a Nikon SMZ1500 stereomicroscope with an HR Plan Apo 1.6x objective or a Deltavision Core imaging system (Applied Precision) with UApo/340 40x 1.35NA objective (Olympus).

Fluorescence microscopy and image processing

Animals were mounted on 2% agarose pads in M9 buffer (Sulston et al., 1983) with 50-100 mM sodium azide depending on developmental stage, and imaged using a Deltavision Core imaging system (Applied Precision) with UApo/340 40x 1.35NA, PlanApo 60x 1.42NA, and U-PlanApo 100x 1.4NA objectives (Olympus) and CoolSnap HQ2 camera. Images were deconvolved using Softworx (Applied Precision) and maximum-brightness projections were obtained from contiguous optical sections using ImageJ.

PBM experiments and data analysis

PBM experiments were performed on universal "all-10-mer" arrays in 8X60K format (Agilent, AMADID 030236) (Berger et al., 2006; Nakagawa et al., 2013). PBM experiments were performed at 500 nM protein concentration in the standard protein binding reaction mixture, substituting buffer A for PBS (buffer A= 138 mM KGlu, 12 mM $NaHCO_3$, 0.8 mM $MgCl_2$, pH 7.2) in the standard PBM protocol (Berger and Bulyk, 2009). Protein binding was detected with an Alexa488-conjugated anti-GST antibody (Life Technologies A-11131), and arrays were scanned using a GenePix 4400A (Molecular Devices) microarray scanner. Binding was quantified using the Universal PBM Analysis Suite (Berger and Bulyk, 2009) to generate E-scores for each 8-mer. Motifs were derived using the Seed-and-Wobble algorithm (Berger and Bulyk, 2009; Berger et al., 2006). Two replicate experiments were performed, with replicate 1 having higher E-scores overall. Replicate 1 is shown in the main body figures, and Replicate 2 is shown in supplemental figures.

Boxplots were generated in R, from the E-scores of the 8-mer sequences that match the UNC-130 FkhP motif ([AG][CT]AAACA) or the FkhS motif

(AA[CT]AACA). Individual data points are displayed on the boxplots using the stripchart function in R. Significant differences in binding were evaluated using a one-sided Mann-Whitney test, with the wilcox.test function in R. All PBM raw data will be available on UniPROBE: http://thebrain.bwh.harvard.edu/uniprobe/ with deposition ID MIZ19A.

Luciferase assays

HEK293T cells were cultured at 37°C and transfected with Fugene (Roche). 48 hours post transfection, cells were collected in cold 1X PBS and transferred into 96-well plates. Renilla and firefly luciferase activity were assayed according to manufacturer's instructions using the Dual-Glo assay (Promega) and bioluminescence was collected on a Molecular Devices Spectramax Paradigm plate reader. Firefly luciferase activity was normalized to renilla luciferase activity in each sample.

REFERENCES

Bacaj, T., Tevlin, M., Lu, Y., and Shaham, S. (2008). Glia are essential for sensory organ function in C. elegans. Science *322*, 744–747.

Berger, M.F., and Bulyk, M.L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. *4*, 393–411.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. *24*, 1429–1435.

Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71–94.

Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science *360*.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502.

Chan, M.M., Smith, Z.D., Grosswendt, S., Kretzmer, H., Norman, T.M., Adamson, B., Jost, M., Quinn, J.J., Yang, D., Jones, M.G., et al. (2019). Molecular recording of mammalian embryogenesis. Nature *570*, 77–82.

Chen, A.I., de Nooij, J.C., and Jessell, T.M. (2006). Graded activity of transcription factor Runx3 specifies the laminar termination pattern of sensory axons in the developing spinal cord. Neuron *49*, 395–408.

Doitsidou, M., Poole, R.J., Sarin, S., Bigelow, H., and Hobert, O. (2010). C. elegans mutant identification with a one-step whole-genome-sequencing and SNP mapping strategy. PloS One *5*, e15435.

Dottori, M., Gross, M.K., Labosky, P., and Goulding, M. (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. Dev. Camb. Engl. *128*, 4127–4138.

Dykes, I.M., Lanier, J., Eng, S.R., and Turner, E.E. (2010). Brn3a regulates neuronal subtype specification in the trigeminal ganglion by promoting Runx expression during sensory differentiation. Neural Develop. *5*, 3.

Eng, S.R., Lanier, J., Fedtsova, N., and Turner, E.E. (2004). Coordinated regulation of gene expression by Brn3a in developing sensory ganglia. Dev. Camb. Engl. *131*, 3859–3870.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360*.

Hammond, T.R., Dufort, C., Dissing-Olesen, L., Giera, S., Young, A., Wysoker, A., Walker, A.J., Gergits, F., Segel, M., Nemesh, J., et al. (2019). Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. Immunity *50*, 253-271.e6.

Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. Curr. Opin. Neurobiol. *56*, 97–105.

Jin, C., Marsden, I., Chen, X., and Liao, X. (1999). Dynamic DNA contacts observed in the NMR structure of winged helix protein-DNA complex. J. Mol. Biol. *289*, 683–690.

John Lin, C.-C., Yu, K., Hatcher, A., Huang, T.-W., Lee, H.K., Carlson, J., Weston, M.C., Chen, F., Zhang, Y., Zhu, W., et al. (2017). Identification of diverse astrocyte populations and their malignant analogs. Nat. Neurosci. *20*, 396–405.

Keyte, A., and Hutson, M.R. (2012). The neural crest in cardiac congenital anomalies. Differ. Res. Biol. Divers. *84*, 25–40.

Kos, R., Reedy, M.V., Johnson, R.L., and Erickson, C.A. (2001). The winged-helix transcription factor FoxD3 is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. Dev. Camb. Engl. *128*, 1467–1479.

Kramer, I., Sigrist, M., de Nooij, J.C., Taniuchi, I., Jessell, T.M., and Arber, S. (2006). A role for Runx transcription factor signaling in dorsal root ganglion sensory neuron diversification. Neuron *49*, 379–393.

Krishnakumar, R., Chen, A.F., Pantovich, M.G., Danial, M., Parchem, R.J., Labosky, P.A., and Blelloch, R. (2016). FOXD3 Regulates Pluripotent Stem Cell Potential by Simultaneously Initiating and Repressing Enhancer Activity. Cell Stem Cell *18*, 104–117.

Levanon, D., Bettoun, D., Harris-Cerruti, C., Woolf, E., Negreanu, V., Eilam, R., Bernstein, Y., Goldenberg, D., Xiao, C., Fliegauf, M., et al. (2002). The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. EMBO J. *21*, 3454–3463.

Lister, J.A., Cooper, C., Nguyen, K., Modrell, M., Grant, K., and Raible, D.W. (2006). Zebrafish Foxd3 is required for development of a subset of neural crest derivatives. Dev. Biol. *290*, 92–104.

Liu, X., Chen, W., Li, W., Li, Y., Priest, J.R., Zhou, B., Wang, J., and Zhou, Z. (2019). Single-Cell RNA-Seq of the Developing Cardiac Outflow Tract Reveals Convergent Development of the Vascular Smooth Muscle Cells. Cell Rep. *28*, 1346-1361.e4.

Lukoseviciute, M., Gavriouchkina, D., Williams, R.M., Hochgreb-Hagele, T., Senanayake, U., Chong-Morrison, V., Thongjuea, S., Repapi, E., Mead, A., and Sauka-Spengler, T. (2018). From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. Dev. Cell *47*, 608-628.e6.

Marmigère, F., Montelius, A., Wegner, M., Groner, Y., Reichardt, L.F., and Ernfors, P. (2006). The Runx1/AML1 transcription factor selectively regulates development and survival of TrkA nociceptive sensory neurons. Nat. Neurosci. *9*, 180–187.

Marques, S., Zeisel, A., Codeluppi, S., van Bruggen, D., Mendanha Falcão, A., Xiao, L., Li, H., Häring, M., Hochgerner, H., Romanov, R.A., et al. (2016). Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. Science *352*, 1326–1329.

Marques, S., van Bruggen, D., Vanichkina, D.P., Floriddia, E.M., Munguba, H., Väremo, L., Giacomello, S., Falcão, A.M., Meijer, M., Björklund, Å.K., et al. (2018). Transcriptional Convergence of Oligodendrocyte Lineage Progenitors during Development. Dev. Cell *46*, 504-517.e7.

Masuda, T., Sankowski, R., Staszewski, O., Böttcher, C., Amann, L., Sagar, null, Scheiwe, C., Nessler, S., Kunz, P., van Loo, G., et al. (2019). Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. Nature *566*, 388–392.

McKenna, A., and Gagnon, J.A. (2019). Recording development with single cell dynamic lineage tracing. Dev. Camb. Engl. *146*.

Mello, C., and Fire, A. (1995). DNA transformation. Methods Cell Biol. *48*, 451–482.

Minevich, G., Park, D.S., Blankenberg, D., Poole, R.J., and Hobert, O. (2012). CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. Genetics *192*, 1249–1269.

Morel, L., Chiang, M.S.R., Higashimori, H., Shoneye, T., Iyer, L.K., Yelick, J., Tai, A., and Yang, Y. (2017). Molecular and Functional Properties of Regional Astrocytes in the Adult Brain. J. Neurosci. Off. J. Soc. Neurosci. *37*, 8706–8717.

Murray, J.I., Boyle, T.J., Preston, E., Vafeados, D., Mericle, B., Weisdepp, P., Zhao, Z., Bao, Z., Boeck, M., and Waterston, R.H. (2012). Multidimensional regulation of gene expression in the C. elegans embryo. Genome Res. *22*, 1282–1294.

Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L., and Bulyk, M.L. (2013). DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc. Natl. Acad. Sci. U. S. A. *110*, 12349–12354.

Nash, B., Colavita, A., Zheng, H., Roy, P.J., and Culotti, J.G. (2000). The forkhead transcription factor UNC-130 is required for the graded spatial expression of the UNC-129 TGF-beta guidance factor in C. elegans. Genes Dev. *14*, 2486–2500.

Ono, H., Kozmik, Z., Yu, J.-K., and Wada, H. (2014). A novel N-terminal motif is responsible for the evolution of neural crest-specific gene-regulatory activity in vertebrate FoxD3. Dev. Biol. *385*, 396–404.

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., et al. (2019). A lineage-resolved molecular atlas of C. elegans embryogenesis at single cell resolution. BioRxiv 565549.

Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495.

Raj, B., Gagnon, J.A., and Schier, A.F. (2018). Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT. Nat. Protoc. *13*, 2685–2713.

Sammut, M., Cook, S.J., Nguyen, K.C.Q., Felton, T., Hall, D.H., Emmons, S.W., Poole, R.J., and Barrios, A. (2015). Glia-derived neurons are required for sex-specific learning in C. elegans. Nature *526*, 385–390.

Sarafi-Reinach, T.R., and Sengupta, P. (2000). The forkhead domain gene unc-130 generates chemosensory neuron diversity in C. elegans. Genes Dev. *14*, 2472–2485.

Sasai, N., Mizuseki, K., and Sasai, Y. (2001). Requirement of FoxD3-class signaling for neural crest determination in Xenopus. Dev. Camb. Engl. *128*, 2525–2536.

Spitzer, S.O., Sitnikov, S., Kamen, Y., Evans, K.A., Kronenberg-Versteeg, D., Dietmann, S., de Faria, O., Agathou, S., and Káradóttir, R.T. (2019). Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age. Neuron *101*, 459-471.e5.

Stewart, R.A., Arduini, B.L., Berghmans, S., George, R.E., Kanki, J.P., Henion, P.D., and Look, A.T. (2006). Zebrafish foxd3 is selectively required for neural crest specification, migration and survival. Dev. Biol. *292*, 174–188.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Teng, L., Mundell, N.A., Frist, A.Y., Wang, Q., and Labosky, P.A. (2008). Requirement for Foxd3 in the maintenance of neural crest progenitors. Dev. Camb. Engl. *135*, 1615–1624.

Wagner, D.E., Weinreb, C., Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science *360*, 981–987.

Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H., and Plasterk, R.H. (2001). Rapid gene mapping in Caenorhabditis elegans using a high density polymorphism map. Nat. Genet. *28*, 160–164.

Yaklichkin, S., Steiner, A.B., Lu, Q., and Kessler, D.S. (2007a). FoxD3 and Grg4 physically interact to repress transcription and induce mesoderm in Xenopus. J. Biol. Chem. *282*, 2548–2557.

Yaklichkin, S., Vekker, A., Stayrook, S., Lewis, M., and Kessler, D.S. (2007b). Prevalence of the EH1 Groucho interaction motif in the metazoan Fox family of transcriptional regulators. BMC Genomics *8*, 201.

Yoshimura, S., Murray, J.I., Lu, Y., Waterston, R.H., and Shaham, S. (2008). mls-2 and vab-3 Control glia development, hlh-17/Olig expression and glia-dependent neurite extension in C. elegans. Dev. Camb. Engl. *135*, 2263–2275.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999-1014.e22.

Figure 1

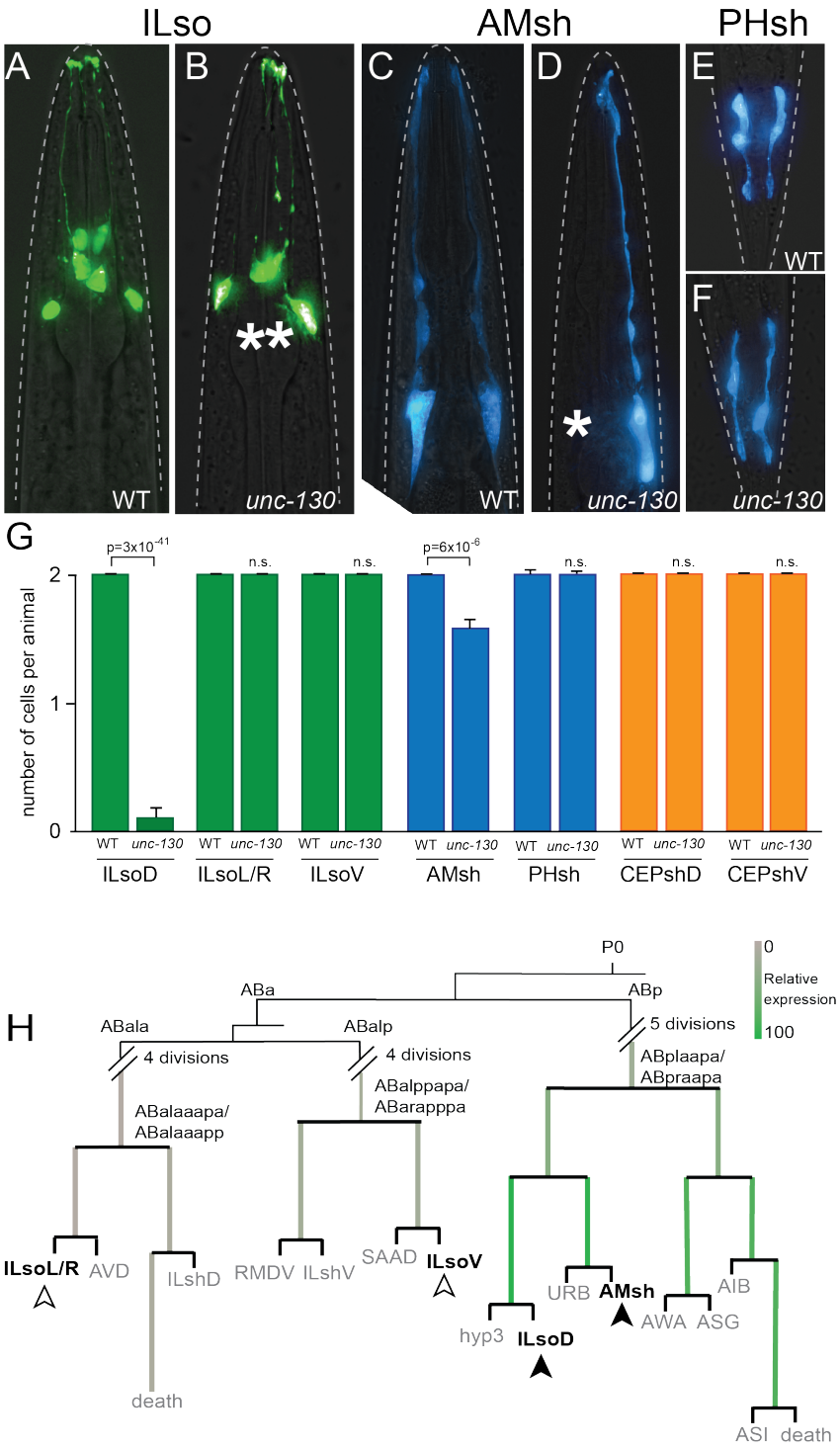
ILso
AMsh
PHsh
A
B
C
D
E
F
WT
unc-130
WT
unc-130
WT
unc-130
G
number of cells per animal
p=3x10^-41
p=6x10^-6
n.s.
WT
unc-130
ILsoD
ILsoL/R
ILsoV
AMsh
PHsh
CEPshD
CEPshV
H
P0
ABa
ABp
ABala
ABalp
4 divisions
5 divisions
ABalaaapa/
ABalaaapp
ABalppapa/
ABarapppa
ABplaapa/
ABpraapa
Relative
expression
0
100
ILsoL/R
AVD
ILshD
RMDV
ILshV
SAAD
ILsoV
URB
AMsh
AIB
AWA
ASG
hyp3
ILsoD
death
ASI
death

Fig. 1. *unc-130* is a lineage-specific regulator of glial fates

Wild-type (A) and *unc-130* mutant (B) animals expressing ILso marker, *grl-18*pro:YFP. Wild-type (C, E) and *unc-130* mutant (D, F) animals expressing AM and PH sheath marker, *F16F9.3*pro:mCherry. Asterisks denote missing cells. (G) Average number of cells expressing cell type-specific marker per animal. ILso – *grl-18*pro; AMsh and PHsh - *F16F9.3*pro; CEPsh - *hlh-17*pro. n > 50 animals per genotype. Error bars – SEM; p-values calculated by Welch's t-test. (H) Lineage diagram of a subset of embryonic cell divisions derived from the AB blastomere that give rise to neurons, glia, and hypodermal cells in the head. Relative expression of *unc-130* is marked by green lines (adapted from (Murray et al., 2012)).

Figure 2

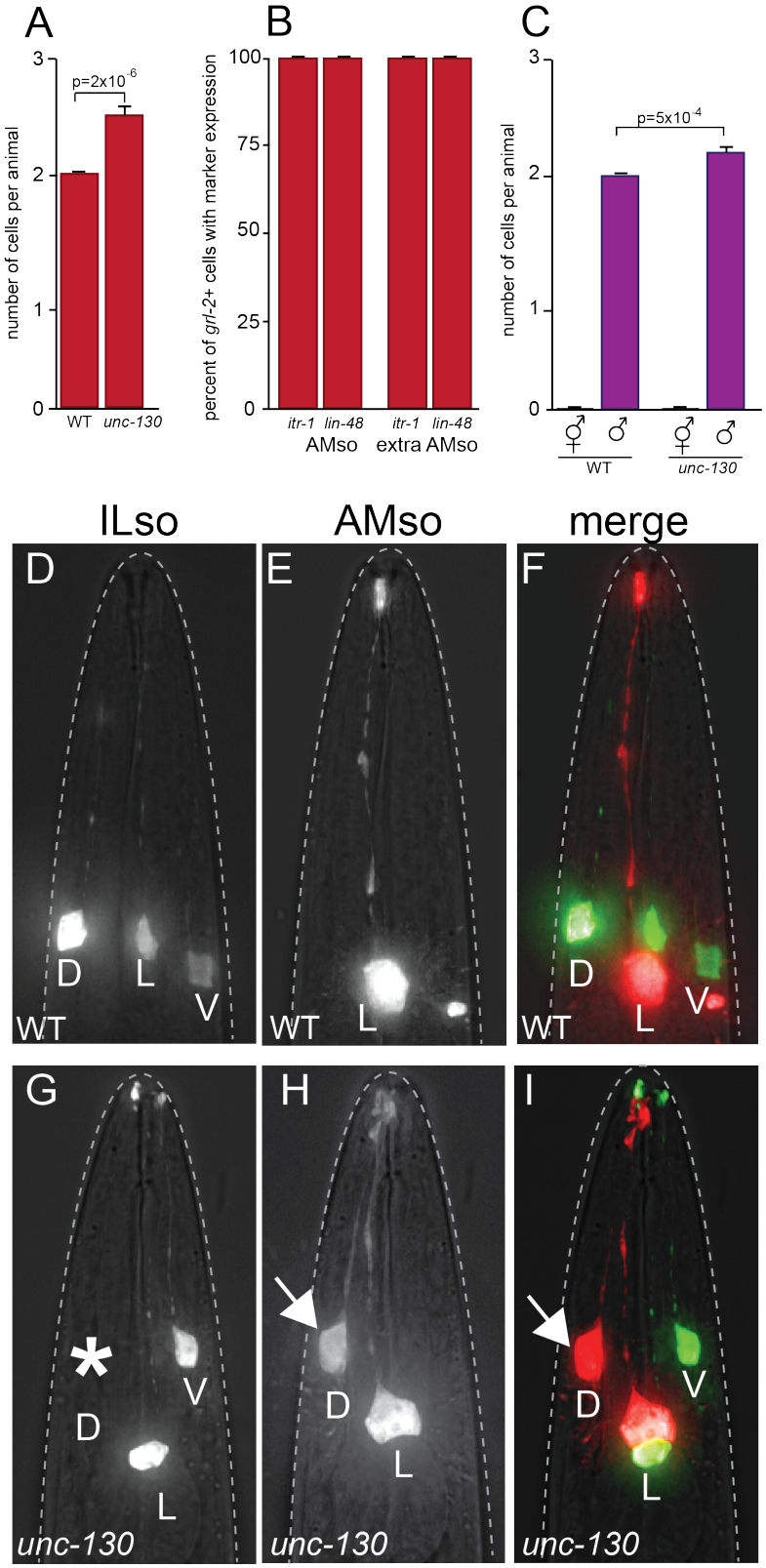
A
B
C
p=2x10-6
number of cells per animal
WT
unc-130
percent of grl-2+ cells with marker expression
itr-1
lin-48
AMso
extra AMso
p=5x10-4
♀
♂
ILso
AMso
merge
D
E
F
G
H
I
WT
unc-130
*

Fig. 2. Extra AMso glia appear in place of missing ILsoD glia

(A) Average number of cells labeled by *grl-2* expression in wild-type and *unc-130* mutant animals. (B) Average number of cells labeled by MCM marker, *pdf-1*pro:RFP, in wild-type and *unc-130* mutant hermaphrodites and males. n = 50 animals per genotype. Error bars – SEM; p-values calculated by Welch's t-test. (C) Percentage of endogenous and extra AMso *grl-2*+ cells that co-express AMso markers, *lin-48* and *itr-1,* in wild-type and *unc-130* mutant animals. n ≥ 20 cells. Lateral, partial volume views of wild-type animals co-expressing ILso marker, *grl-18*pro:YFP (D) and AMso marker, *grl-2*pro:YFP (E) and merge (F), showing dorsal-ventral positioning of cells in the head. Lateral, partial volume views of *unc-130* mutant animals co-expressing ILso marker, *grl-18*pro:YFP (G) and AMso marker, *grl-2*pro:YFP (H) and merge (I), showing dorsal-ventral positioning of cells in the head. Asterisk, missing ILsoD. Arrow, extra AMso. D-dorsal, L-lateral, V-ventral.

Figure 3

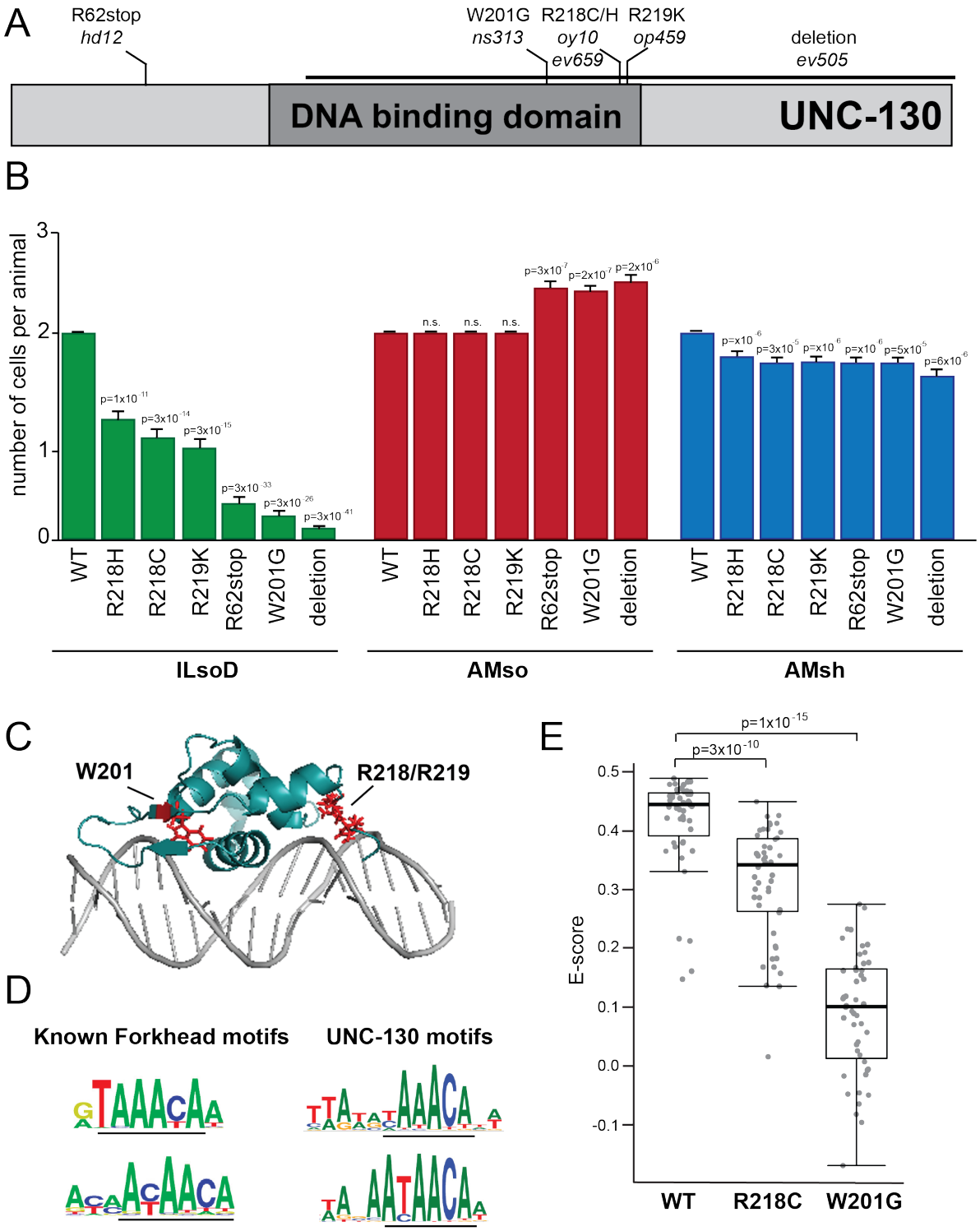
A
R62stop
hd12
W201G
ns313
R218C/H
oy10
ev659
R219K
op459
deletion
ev505
DNA binding domain
UNC-130
B
number of cells per animal
p=1x10-11
p=3x10-14
p=3x10-15
p=3x10-33
p=3x10-26
p=3x10-41
n.s.
p=3x10-7
p=2x10-7
p=2x10-6
p=x10-6
p=3x10-5
p=x10-6
p=x10-6
p=5x10-5
p=6x10-6
WT
R218H
R218C
R219K
R62stop
W201G
deletion
ILsoD
AMso
AMsh
C
W201
R218/R219
D
Known Forkhead motifs
UNC-130 motifs
E
p=1x10-15
p=3x10-10
E-score
WT
R218C
W201G

Fig. 3. Severity of glial phenotypes correlate with UNC-130:DNA binding defects

(A) Schematic diagram of UNC-130 protein with DNA binding domain (green) highlighted. Location of point mutations and deletion are indicated. (B) Average number of ILsoD glia (black, *grl-18*pro:GFP), AMso glia (dark gray, *grl-2*pro:YFP), and AMsh glia (light gray, *F16F9.3*pro:mCherry) per animal in all genotypes. n = 50 animals per genotype. Error bars – SEM; p-values calculated by Welch's t-test. (C) Structure of FoxD3 DNA binding domain (green) interacting with DNA (gray) (Protein Data Bank ID code 2HDC, (Jin et al., 1999)). W201, R218, and R219 residues are highlighted in red. (D) Logos of vertebrate primary (top) and secondary (bottom) Forkhead motifs (Nakagawa et al., 2013) and logos of UNC-130 preferred DNA sequences that resemble primary (top) and secondary (bottom) motifs as determined by this study. Core motifs are underlined. (E) Scatter plot of E-scores for 8-mer DNA sequences matching [A/G][C/T]AAACA or AA[C/T]AACA from protein binding microarray assays of wild-type, R218C, and W201G mutant UNC-130 proteins. Black lines represent population median; top and bottom of boxes are 25th and 75th percentiles, respectively; and top and bottom of whiskers are either most extreme point or 1.5x the interquartile range. p-values calculated by Mann-Whitney test.

Figure 4

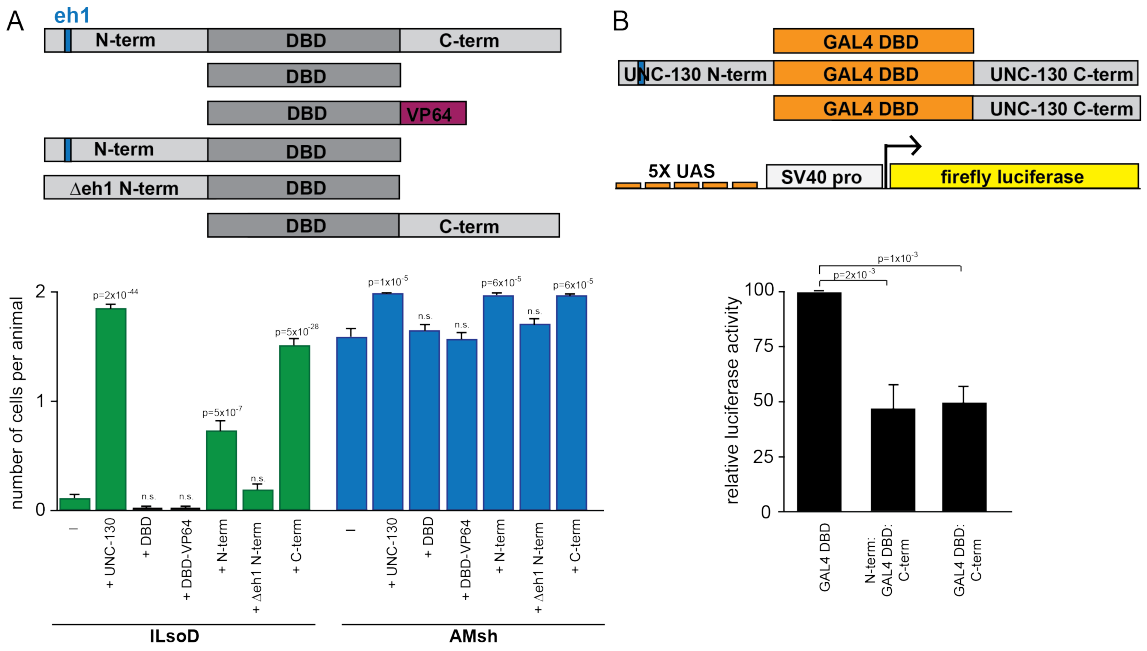

Fig. 4. UNC-130 acts as a repressor to specify glial fates

(A) Schematic diagram of UNC-130 constructs used in rescue experiments to determine domain function (top). *unc-130* promoter was used to drive expression of individual constructs in the *ev505* deletion mutant strain and extent of rescue was assessed. Average number of ILsoD (black) and AMsh (light gray) cells present in each condition (bottom). n = 50 animals per genotype. Error bars – SEM; p-values calculated by Welch's t-test. (B) Schematic diagram of constructs used in luciferase assays (top). GAL4DBD alone, N-term:GAL4DBD:C-term, or GAL4DBD:C-term, along with a UAS-firefly luciferase reporter and constitutively expressed renilla luciferase were transfected into HEK293T cells. Relative firefly luciferase activity, first normalized to renilla luciferase bioluminescence in each sample, and then to DBD-GAL4 alone (bottom). Error bars - SD; p-values calculated by Welch's t-test.

Figure 5

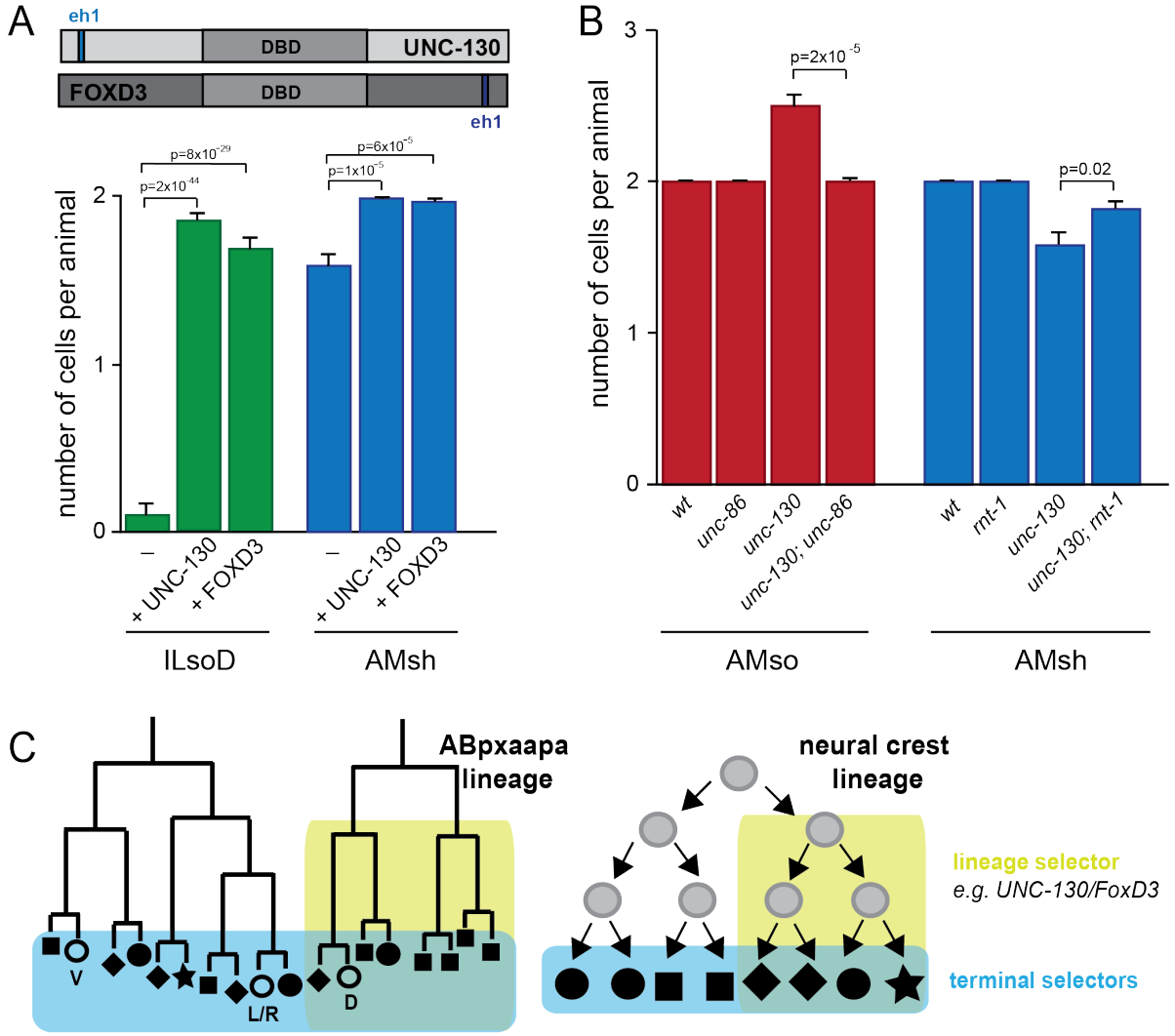
A
eh1
DBD
UNC-130
FOXD3
DBD
eh1
number of cells per animal
p=8x10^-29
p=2x10^-44
p=6x10^-5
p=1x10^-5
–
+ UNC-130
+ FOXD3
–
+ UNC-130
+ FOXD3
ILsoD
AMsh
B
number of cells per animal
p=2x10^-5
p=0.02
wt
unc-86
unc-130
unc-130; unc-86
wt
rnt-1
unc-130
unc-130; rnt-1
AMso
AMsh
C
ABpxaapa lineage
neural crest lineage
V
L/R
D
lineage selector
e.g. UNC-130/FoxD3
terminal selectors

Fig. 5. UNC-130 acts in conserved pathways to specify glial fate

(A) Schematic diagram of UNC-130 and FOXD3 proteins. DNA-binding domain – green. eh1 motif – blue (top). UNC-130 or human FOXD3 coding sequence were expressed under the *unc-130* promoter in the *ev505* deletion mutant strain and extent of rescue was assessed. Average number of ILsoD cells (black) and AMsh cells (light gray) present in each condition (bottom). (B) Average number of AMso cells per animal (dark gray) in *unc-130* and *unc-86* single and double mutants. Average number of AMsh cells per animal (light gray) in *unc-130* and *rnt-1* single and double mutants. n = 50 animals per genotype. Error bars – SEM; p-values calculated by Welch's t-test. (C) Lineage selector model. UNC-130 and FoxD3 act throughout a sublineage to affect diverse cell types (yellow) whereas terminal selectors act in differentiating cells to affect a single cell type (blue).