TRANSPARENT PROCESS

OPEN ACCESS

Single-cell RNA sequencing of motoneurons identifies regulators of synaptic wiring in *Drosophila* embryos

Jessica Velten[1,2,3], Xuefan Gao[1] iD, Patrick Van Nierop y Sanchez[1] iD, Katrin Domsch[1,4], Rashi Agarwal[1] iD, Lena Bognar[1], Malte Paulsen[3], Lars Velten[2,5,*] iD & Ingrid Lohmann[1,**] iD

Abstract

The correct wiring of neuronal circuits is one of the most complex processes in development, since axons form highly specific connections out of a vast number of possibilities. Circuit structure is genetically determined in vertebrates and invertebrates, but the mechanisms guiding each axon to precisely innervate a unique pre-specified target cell are poorly understood. We investigated *Drosophila* embryonic motoneurons using single-cell genomics, imaging, and genetics. We show that a cell-specific combination of homeodomain transcription factors and downstream immunoglobulin domain proteins is expressed in individual cells and plays an important role in determining cell-specific connections between differentiated motoneurons and target muscles. We provide genetic evidence for a functional role of five homeodomain transcription factors and four immunoglobulins in the neuromuscular wiring. Knockdown and ectopic expression of these homeodomain transcription factors induces cell-specific synaptic wiring defects that are partly phenocopied by genetic modulations of their immunoglobulin targets. Taken together, our data suggest that homeodomain transcription factor and immunoglobulin molecule expression could be directly linked and function as a crucial determinant of neuronal circuit structure.

Introduction

Neuronal circuits in mammals as well as *Drosophila* are stereotypically wired for the precise execution of functional tasks critical for organismal survival. The formation of such circuits is a step-wise process, which starts with the specification of neuronal cell types and their accurate arrangements in space, followed by the correct wiring of individual cells and their final integration into a functional network. To ensure such precision, the structure and connectivity of neural circuits is genetically specified. However, how these complex interconnected processes are encoded in the genome and executed by the cellular protein machinery is still not fully understood.

According to the "labelled pathway hypothesis" (Sperry, 1963), neurons stochastically and transiently form contacts with many possible targets after their specification, while the expression of specific cell surface proteins (CSPs) is thought to stabilize the correct connections, a process called synaptic specificity (Sanes & Zipursky, 2020). Many lines of evidence support this hypothesis. Recent work has shown that combinations of immunoglobulin superfamily (IgSF) cell surface proteins (Dprs) are differentially expressed in distinct neuronal clusters and bind to specific Dpr binding proteins (DIPs) expressed in synaptic partners (Nakamura *et al*, 2002; Özkan *et al*, 2013; Carrillo *et al*, 2015). In the visual system, combinations of CSPs are differentially expressed between layers (Tan *et al*, 2015), while in olfactory neurons, a combinatorial expression of transcription factors (TFs) and CSPs maps neurons with the same olfactory receptor to the same glomerulus (Couto *et al*, 2005; Li *et al*, 2017, 2020; McLaughlin *et al*, 2021). All these studies explain how groups of similar neuronal cells are molecularly defined and provide a hypothesis on how stereotypic connections to another neuronal cell type are formed. By contrast, how the specificity of circuits is specified and controlled at the level of single cells is still not completely resolved.

In the *Drosophila* neuromuscular system, every single motoneuron (MN) forms unique and stereotypic connections with target muscles already during embryogenesis, but the molecular mechanisms underlying this specificity is unclear (Allan & Thor, 2015). MNs are progressively specified from anterior to posterior by segment-specific TFs (Bossing *et al*, 1996; Schmidt *et al*, 1997;

---

1 Department of Developmental Biology, Centre for Organismal Studies (COS) Heidelberg, Heidelberg, Germany
2 The Barcelona Institute of Science and Technology, Centre for Genomic Regulation (CRG), Barcelona, Spain
3 Flow Cytometry Core Facility, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
4 Developmental Biology, Erlangen-Nürnberg University, Erlangen, Germany
5 Universitat Pompeu Fabra (UPF), Barcelona, Spain
*Corresponding author. Tel: +34 93 316 0185; E-mail: lars.velten@crg.eu
**Corresponding author. Tel: +49 6221 545523; Fax: +49 6221 546424; E-mail: ingrid.lohmann@cos.uni-heidelberg.de

Angelini & Kaufman, 2005) and further along the dorsal to ventral axis (Broihier & Skeath, 2002; Broihier *et al*, 2004; Landgraf & Thor, 2006) before extending their nerve projections to predefined locations specified in all three spatial dimensions (Landgraf *et al*, 1997; Thor *et al*, 1999; Broihier & Skeath, 2002; Broihier *et al*, 2004; Zarin *et al*, 2014; Hessinger *et al*, 2016). These observations have suggested region-specific mechanisms in the determination of connectivity patterns. However, such a regional model alone is unlikely to explain the precise connectivity patterns of single cells (Nassif *et al*, 1998; Landgraf *et al*, 2003). In addition, elegant studies in the vertebrate central nervous system and classical transplantation experiments demonstrated that positional identity and connectivity patterns of single neurons are stably maintained even after experimental relocation of cells (Demireva *et al*, 2011). Thus, there seems to be a molecular mechanism that stably imprints cellular identity and instructs the formation of cellular connectivity.

Homeodomain TFs have long been known to play important roles in the specification, differentiation and maintenance of neurons, also in MNs, in different organisms (Thor *et al*, 1999; Thor & Thomas, 2002; Urbach *et al*, 2006, 2016; Sanguinetto *et al*, 2008; Philippidou *et al*, 2012; Deneris & Hobert, 2014; Allan & Thor, 2015; Zeisel *et al*, 2018; Domsch *et al*, 2019; Sugino *et al*, 2019; Allen *et al*, 2020; Reilly *et al*, 2020). Importantly, it has been shown just recently that each neuron class in the nematode *C. elegans* expresses a unique combination of homeodomain TFs, which is unambiguously associated with neuronal identities (Reilly *et al*, 2020; Hobert, 2021). However, it is so far unclear whether this concept extends to other organisms; whether such cell-specific combinations of homeodomain TFs also instruct later events in circuit formation; and, finally, which molecules downstream of such homeodomain TFs realize synaptic target choice and specificity at the single-cell level.

Using single-cell RNA sequencing (scRNA-Seq) with high numbers of biological replicates, we demonstrate that cell-specific expression of multiple homeodomain TFs is associated with the cellular heterogeneity within differentiated MNs along the major body axes of *Drosophila* embryos. We furthermore show that multiple CSPs, in particular cell surface immunoglobulins (Igs), act downstream of homeodomain TFs in individual MNs and play an important role in determining specificity during the synaptic wiring phase. Knockdown and ectopic expression of homeodomain TFs induces synaptic wiring defects specific to single cells that are partly phenocopied by genetic manipulation of their putative Ig targets. Additionally, our data suggest that shared combinations of homeodomain TF are expressed in matching synaptic partners of functional neuronal circuits. Based on these findings, we propose that the development of individual neuronal circuits is genetically defined by a linked "homeo-immunoglobulin program", which serves as one of the major determinants for complex neuronal wiring with single cell precision.

Results

A reference map of MNs during the synaptic wiring phase in *Drosophila* embryos

We aimed at identifying molecules driving specificity in synaptic wiring at the single-cell level to gain a comprehensive view of the complex yet highly precise synaptic matching process. To this end, we used the *Drosophila* neuromuscular system as our model, as it is ideally suited to study mechanisms of synaptic specificity: first, this system is of relatively low complexity; and second, it is fully established at the end of embryogenesis with about 30-35 MNs innervating in a highly stereotypic manner 30 muscles in each abdominal hemisegment of stage 17 embryos (Landgraf *et al*, 1997; Hoang & Chiba, 2001; Landgraf & Thor, 2006; Kim *et al*, 2009; Couton *et al*, 2015). Evidently, selection of the proper developmental stage is critical for the comprehensive identification of cues driving the highly specific interaction of neuronal cells. This notion is based on previous studies showing that neurons diversify most on the transcriptional level when they are in the process of contacting their synaptic partners while their transcriptomes become indistinguishable upon completion of neuronal connectivity (Li *et al*, 2017). In the *Drosophila* neuromuscular system, embryonic MNs interact with their muscle partners at the end of embryonic stage 16 (Landgraf *et al*, 1997), suggesting motoneuronal transcriptomes to be most diverse at this developmental stage. Based on these considerations, we performed scRNA-Seq of stage 16 embryonic cells marked by the *OK371*-GAL4 driver (Mahr & Aberle, 2006) controlling the expression of the UAS-*RFP* transgene (Fig EV1A and B). This driver is based on a regulatory element controlling the expression of the presynaptic vesicular glutamate transporter (VGlut) and is active specifically in all/most MNs at late stages of embryogenesis (stage 17) (Mahr & Aberle, 2006) (Fig EV1A). In addition, the *OK371*-GAL4 driver is active in a few glutamatergic brain neurons (Fig EV1A), which were later excluded by restricting the analysis to *Hox*-expressing cells. We further confirmed that known motoneuronal subtypes are targeted by this driver in expected ratios in stage 16 embryos (Fig EV3A–D, and see below) (Mahr & Aberle, 2006).

For the experiment, single *RFP*-expressing MNs were sorted from a pool of precisely staged embryos by fluorescence-activated cell sorting (FACS) (Fig 1A). In total, 1,536 MNs were sequenced by SMART-Seq2 (Picelli *et al*, 2014) from pooled embryos. After filtering based on a minimum of 500 genes observed with 10 reads each, 999 single-cell transcriptomes were retained (Fig EV1D). Thus, every biologically unique motoneuronal cell (~140 *OK371*-positive cells in a single embryo, Fig EV1A–C) was sequenced in approximately 7 biological replicates in our dataset. By comparing our dataset to recently published data, we confirmed that the quality of our data matches the standards in the field, in particular with regard to sequencing depth (Fig EV1E and F). A median of 1,202 unique genes were observed per cell, and a negligible fraction of 0.1% of reads mapped to the mitochondrial genome, supporting the high technical quality of the data. Abundant expression of motoneuronal marker genes like *Vesicular glutamate transporter (VGlut)* and *embryonic lethal abnormal vision (elav)* indicated successful sorting of the targeted cell population (Fig EV2B).

Hox genes are known to be expressed in a consecutive order along the AP axis of *Drosophila*. We used this property to precisely locate single-cell transcriptomes along the AP axis as further described below. To this end, we implemented a custom modification of the SMART-Seq2 protocol by adding primers targeting each *Hox* gene to the reverse transcription (RT) and preamplification step that permits an increased representation of the lowly expressed *Hox* genes as spatial markers (Fig 1A, see Materials and Methods) (Giustacchini *et al*, 2017; Velten *et al*, 2021). Despite the low expression
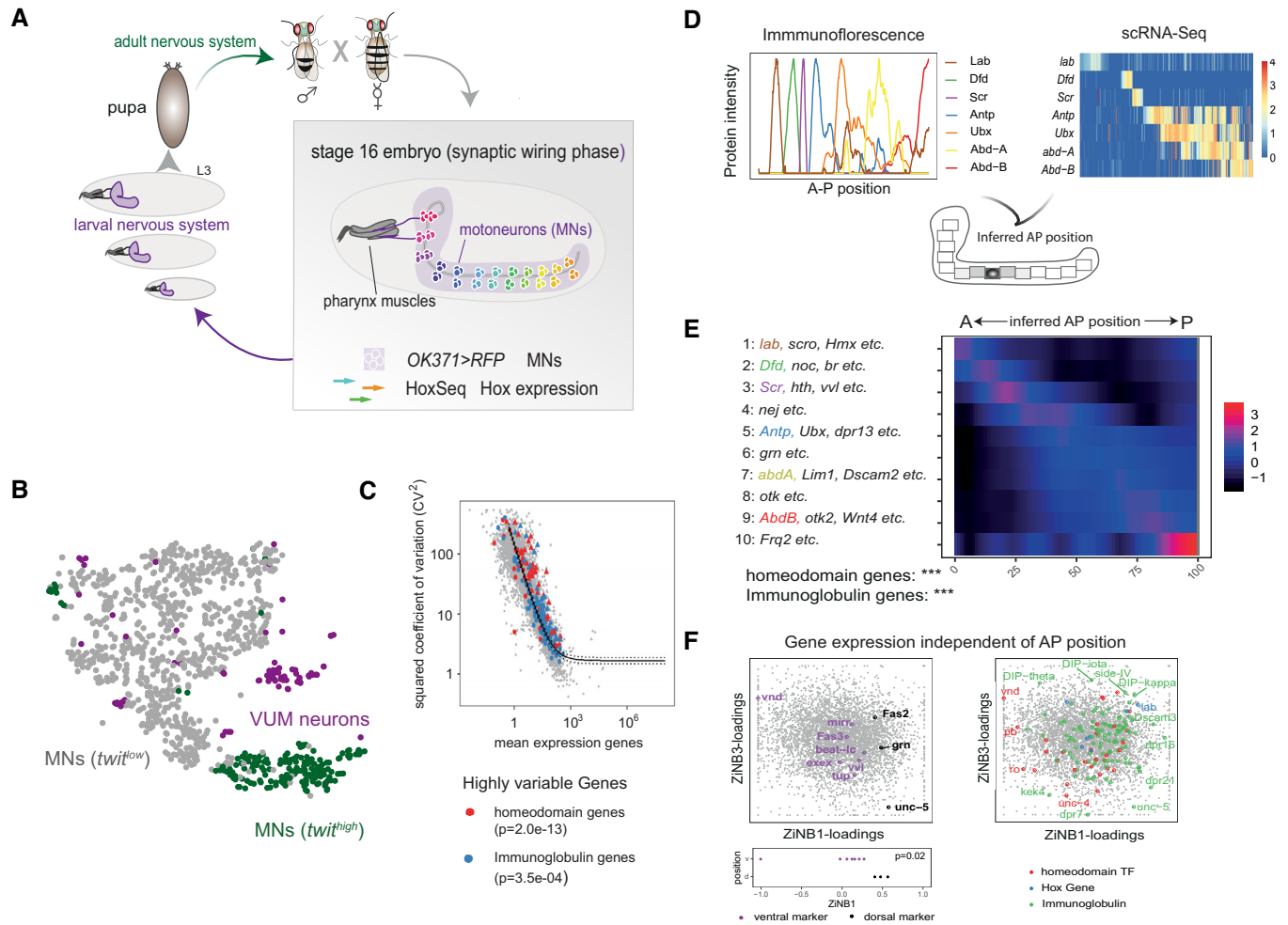
Figure 1. scRNA-Seq identifies highly variable homeodomain TF expression in late embryonic MNs at the time of synaptic wiring.

A Schematic drawing depicts the step-wise development of the nervous system in *Drosophila*, starting during embryogenesis, progressing through three larval stages and metamorphosing during pupal stage into the adult nervous system. The first connections in the neuromuscular system are formed between MNs and muscles in late *Drosophila* embryos (stage 16). MNs at this stage expressing UAS-*RFP* under the control of the motoneuronal driver *OK371*-GAL4 (*OK371 > RFP*) are color-coded along the ventral nerve cord according to different patterns of *Hox* gene expression. *OK371-RFP*-positive MNs were FACS sorted and single cells were sequenced by targeted Smart-Seq2 to enrich for *Hox* gene representation (HoxSeq) as spatial markers (see also Materials and Methods).

B t-distributed stochastic neighbour embedding (t-SNE) plot of *n* = 999 single-cell transcriptomes. Colours correspond to three clusters, VUM neurons (purple), *twit^low^* (grey) and *twit^high^* (green) MNs.

C Identification of highly variable genes in the *twit^low^* cluster using the method by (Brennecke *et al*, 2013). Scatter plot depicts the mean expression for each gene and squared coefficient of variation across *twit^low^* cells. The solid line indicates the fit, dashed lines the 95% confidence interval. Genes with a significantly elevated variance are shown as triangles, other genes as circles. Different gene classes are colour coded. *P*-values shown were retrieved by a hypergeometric test for enrichment of the respective gene class among highly variable genes.

D Outline of strategy to map single MNs to a position along the AP axis (see also Materials and Methods). Upper panel: Intensities of Hox protein expression along the ventral nerve cord measured by immunofluorescence (upper left panel) and co-expression patterns of *Hox* gene transcripts measured by scRNA-Seq (upper right panel) were used as input. Upper right panel depicts a heatmap of color-coded Hox gene expression levels; columns correspond to *n* = 758 single *twit^low^* MNs. Lower panel: AP position is inferred form scRNA-Seq data by probabilistically mapping *Hox* gene expression pattern in each individual cell to the immunofluorescence reference data.

E Genes with significant variation along the AP axis were identified and clustered into 10 groups of distinct expression patterns (Materials and Methods). Heatmap shows average gene expression per cluster (rows) across single cells (columns). Asterisks indicate *P*-value of a hypergeometric test for enrichment of protein domains, ***P* < 0.001.

F *Left panel:* ZINB-WaVE (Risso *et al*, 2018, 2019) was used to statistically separate gene expression variability into parts linked to AP position and parts independent thereof. Scatter plot of ZINB-WaVE loadings separates known dorsal and ventral marker genes on ZINB-WaVE component 1. *Right panel:* Genes encoding homeodomain TFs and genes encoding Ig domain molecules (see colour code) show high loadings on ZINB-WaVE components 1 and 3, demonstrating high variability independent of AP position.

of *Hox* genes in late embryonic stages, which is common to all TF encoding genes, we identified 75% of the MNs to express at least one *Hox* gene (Fig EV2C).

To explore the molecular diversity of the MNs, we performed two independent unsupervised analyses, t-distributed neighbour embedding (tSNE) and hierarchical clustering (Figs 1B and EV2A). Both methods identified a cluster corresponding to modulator neurons (VUMs, 8% of the cells) as well as two large, yet distinct clusters of cells that differ in the expression of the marker genes *rhea* and *target of wit* (*twit*). VUM MNs belong to a very distinct MN subtype expressing a combination of subtype-specific marker genes, *Vesicular monoamine transporter (Vmat*; Fig EV2A and D), *Tyramine β* hydroxylase (*Tbh), diacyl glycerol kinase (dgk)* and the motoneuronal marker *Zn finger homeodomain 1* (*zfh1*) (Stagg *et al*, 2011), which we all identified in the VUM neuron cluster. These type II glutamatergic/octopaminergic MNs exhibit modulator roles in taste responses (Sink & Whitington, 1991; Landgraf *et al*, 1997; Siegler & Jia, 1999; Stagg *et al*, 2011), while the *twit*[low] and *twit*[high] cluster can be assigned to the abundant glutamatergic type I MN class (Hoang & Chiba, 2001; Kim *et al*, 2009). *In situ* hybridization chain reactions (HCR) of late stage embryos localized *twit* transcripts in median and lateral clusters of posteriorly located MNs (Fig EV2D) (Kim & Marqués, 2012), suggesting that those two groups of MNs, the VUMs and *twit*[high] MNs, represent indeed two distinct motoneuronal subtypes with different locations.

Taken together, these analyses showed that the dataset generated in this study was of high quality, consistent with published data and provided an approximately 7-fold cellular coverage of each biologically unique MN at the synaptic wiring stage. In addition, we identified three rather homogenous cluster that can be identified by the expression of *twit*[high], *twit*[low] and *Vmat*, corresponding to type I and type II MNs.

The expression of homeodomain transcription factor and Ig domain genes are highly variable within *twit*[low] MNs

To investigate processes required for MN diversity during the synaptic wiring phase, we focused on the largest cluster of MNs expressing low levels of *twit* (*twit*[low]). This cluster contains the majority of glutamatergic type I MNs that are equally distributed along the ventral nerve cord, rather than specialized subtypes of MNs such as the VUMs or *twit*[high] clusters that are unevenly distributed along this embryonic body axis (Fig EV2D). Although this population in the largest cluster (*twit*[low]) appears rather homogenous, our dataset confirms mutual exclusive expression of known markers for the two major motoneuronal subsets, the dorsally and ventrally projecting MNs (Landgraf *et al*, 1999; Certel & Thor, 2004; Garces & Thor, 2006; Landgraf & Thor, 2006; Technau *et al*, 2014; Zarin *et al*, 2014; Urbach *et al*, 2016; Zarin & Labrador, 2019), while common markers are identified in both subsets including *unc5* (Keleman & Dickson, 2001) (Fig EV3B and D). Importantly, the MN subtypes are represented in our dataset in the ratios expected from literature (Landgraf *et al*, 1999; Garces & Thor, 2006; Zarin *et al*, 2019) (Fig EV3C). Hence, these data suggest that the population profiled here is very similar and representative of MNs present at the end of embryogenesis (stage 17).

Using a statistical test to discover biologically variable genes from scRNA-Seq data (Brennecke *et al*, 2013), we observed that

homeodomain TF and Ig domain encoding genes displayed a very high biological variability within *twit*[low] cells (Fig 1C). Similar results were obtained by principal component analysis (PCA) of the *twit*[low] cells, which demonstrated that the expression of homeodomain TFs and CSPs involved in synaptic matching and axon guidance co-varied (Fig EV3E and F). In particular, the expression of known mediators of synaptic specificity, for instance Dpr protein-encoding genes (Özkan *et al*, 2013; Carrillo *et al*, 2015), was highly variable within the *twit*[low] cluster and co-varied with homeodomain TF gene expression (Fig EV3F).

Taken together, more detailed analysis on the largest type I MN cluster (*twit*[low]) revealed that within this cluster all major motoneuronal subtypes were detected and that homeodomain TF and CSP expression was highly variable and co-varied.

Homeodomain TF combinations are associated with spatial variability

We hypothesized that the high variability of homeodomain TF gene expression within otherwise homogenous MNs is caused by spatial cues. To investigate this possibility, we spatially mapped single MNs along the AP body axis using *Hox* gene expression as spatial reference points. We first created a high-resolution map of Hox protein expression by immunofluorescence, showing that anterior Hox proteins were expressed in clearly defined stripes, whereas expression of posterior Hox proteins was partially overlapping (Figs 1D, and EV4A and B). The same co-expression patterns were observed in our scRNA-Seq data (Figs 1D and EV4C), allowing us to probabilistically map cells from the scRNA-Seq dataset to a position along the AP axis (Fig 1D, see Materials and Methods). We validated this mapping strategy by immunofluorescence. To this end, we used the inferred AP position to estimate the expression pattern of every gene along the AP axis (Fig 1E). We thereby identified candidate genes with differential expression along this axis. Importantly, these candidates were not used for constructing the model. For two such candidates, *Frq1* and *hth*, we compared the predicted expression pattern to immunofluorescence data and observed a high agreement (Fig EV4D). Inferred AP position was significantly correlated with principal components 3 and 4 (Fig EV4E), indicating that AP position profoundly affects the entire transcriptome of each cell.

Interestingly, the above-described heterogeneity of homeodomain encoding genes is aligned with the AP position, but we additionally found more variability of homeodomain encoding genes independent of AP position. To that end, we made use of ZINB-WaVE analysis (Risso *et al*, 2018, 2019). We separated scRNA-Seq data into variability linked to the known covariates (AP position and technical variability) and into processes statistically independent thereof (Fig 1F). On the first component of AP-independent variability, we identified one group of genes known as marker for dorsal–ventral (DV) position (Bhat, 1999; Skeath, 1999; Urbach *et al*, 2006). These genes were ordered according to their localization in the embryo from dorsal to ventral. Again, homeodomain TFs and Ig surface proteins were among the most variable genes on this AP-independent axis of variability. Immunofluorescence experiments of two ventral marker genes, *mirror (mirr)* and *ventral veins lacking (vvl)*, confirmed the predicted DV position (Fig EV4F).

Together, these analyses showed that highly variable homeodomain encoding genes are associated with spatial heterogeneity,

which is in line with previous findings (Bhat, 1999; Skeath, 1999; Urbach *et al*, 2006). Furthermore, using *Hox* gene expression as spatial markers allowed us to map individual MNs along the AP axis.

Cell-specific homeodomain TF combinations delineate *Drosophila* embryonic MN heterogeneity during synaptic wiring

Our analyses suggested that homeodomain encoding genes, unlike other groups of genes, can be used to more precisely depict the cellular heterogeneity of *twit*^low MNs. Thus, we performed unsupervised hierarchical clustering using only the highly variable homeobox genes as input. In addition, we used a normalization strategy that effectively classifies lowly expressed genes as "expressed" or "not expressed", whereas the expression of genes with sufficient sequencing coverage was represented in a more quantitative manner (Fig EV4G, see Materials and Methods). Furthermore, cells not expressing any *Hox* genes were excluded from further analyses, as they might include few glutamatergic brain neurons potentially targeted by *OK371*-GAL4 (Fig EV1A). This approach identified small groups of cells expressing a unique combination of homeodomain-containing genes (Fig 2A). Using this strategy, we could show that these patterns were independent of technical covariates and that clustering is not affected by technical noise common to single-cell transcriptomic data (Fig EV4H).

The identification of defined groups of cells with homogenous expression of homeodomain TFs required a splitting of the dataset into a specific number of clusters. Since a statistical specification of cluster numbers from scRNA-Seq data remain an unresolved issue in the field (Zhu *et al*, 2018; Luecken & Theis, 2019), we used our knowledge that embryos at the exact time point of sequencing harbour around 140 *OK371*-positive MNs differentially distributed along the AP axis (Fig EV1A–C). Due to the bilateral symmetry, this corresponds to 70 biologically unique cellular identities, including the *twit*^high, *twit*^low and VUM cells. We then arbitrarily clustered the data into 60 groups, corresponding to the estimated number of *twit*^low cells (Fig EV2D).

We next validated that these groups correspond to defined cells in a few cases by using the stereotypic position of axon projections as reference for defined cells. As previously described (Friedrich *et al*, 2016), we could reproducibly identify the MN innervating the mouth hook elevator muscle (MHE) by its Dfd expression and FasII staining (Figs 2B and C, and 3A). Based on its similar position in *Calliphora vicina* (Schoofs *et al*, 2010a, 2010b), we termed this axon projection MN2a. The neurons directly adjacent to MN2a were termed MN2c and MN4, while the MN innervating the mouth hook depressor muscle (MHD) was called MN3 (Schoofs *et al*, 2010a, 2010b). MN2a, MN2c, MN3 and MN4 were identified based on their position with regard to the FasII-stained axon projections. We then separately measured protein or RNA expression of Lab, Dfd, Scr, Vvl, Lim1, Hth, *unc-4* and *mirr* in these MNs using immunofluorescence and hybridization chain reaction (HCR; Fig 2D and E). Based on the expression pattern of these eight genes, we were able to assign all four MNs unambiguously to clusters defined by the transcriptome data (Dataset EV1; MN2c = cluster C6, MN2a = cluster C18, MN4 = cluster C12, MN3 = cluster C20).

While difficulties in microscopically identifying all individual 60 MNs impeded us from unanimously designating all clusters as biologically unique cells, our analyses demonstrated that a homeodomain TF combination is associated with cellular heterogeneity during the synaptic wiring phase and is specific to small groups of cells or even single cells.

Homeodomain TFs modulate synaptic target specificity of two distinct MN projections

After having identified cell-specific combinations of homeodomain TFs associated with motoneuronal heterogeneity, we wanted to test their contribution to synaptic specificity in individual embryonic MNs. To this end, we focused our analysis on MN2a, which co-expresses the homeodomain TFs Dfd, Mirr and Hth, and its projection to the MHE muscle (Fig 3A). We used the pan-neural driver *elav*-GAL4 (Luo *et al*, 1994) to interfere with these three TF genes by RNAi-mediated gene silencing (RNAi) and examined synaptic defects of MN2a in stage 17 *Drosophila* embryos. The *elav*-GAL4 driver is ideally suited for temporal interference with proteins during the late synaptic wiring phase, as RNAi-mediated knockdown is realized earliest in stage 14 of motoneuronal axonogenesis (Figs 3B and EV5A), thereby leaving the initial specification of these cells unaffected. To complement this approach, we used additionally a more specific driver for MNs, *OK6*-GAL4, which is active in glutamatergic MNs only (Aberle *et al*, 2002).

Defects on MN2a induced by RNAi were classified into two distinct categories, wiring defects (i.e. mistargeting of neurons to the wrong muscle) and terminal defects (i.e. abnormal synaptic morphologies at axon terminal sites; Fig 3C–E). Terminal defects were quantified but not analysed in detail (Fig 3E). Our experiments showed late neural interference with *Dfd*, *mirr* and *hth* resulted in an abnormal muscle targeting of MN2a (Fig 3D and E); however, in each case, MN2a innervated different ectopic target sites. For example, late neural depletion of *Dfd* led to mistargeting of the MN2a to the labial retractor muscle (LR), a muscle located directly anterior to the MHE muscle, in 29% of the embryos, while interference with *hth* resulted in a mis-guidance of the MN2a to external muscles in 70% of the embryos (Fig 3D and E). In the case of *mirr* knockdown, the MN2a tended to target the MHD muscle located posterior to the MHE; however, wiring defects were less pronounced and frequent (Fig 3D and E). Importantly, defects observed in *elav* > *Dfd*^RNAi and *OK6* > *Dfd*^RNAi embryos were comparable (Figs 3D and 3E, and EV5E), showing that the phenotypes were primarily due to knockdown of gene expression in MNs and not or to a minor extent caused by non-autonomous effects from non-MNs labelled by *elav*-GAL4. In the case of Dfd depletion, we additionally investigated behavioural defects of mouth hook movements by hatching rate assays and measured decreased rates (Fig 3F). Neural interference with one homeodomain TF did not affect the expression of other homeodomain TF genes, which we exemplarily showed for *Lim1* or *unc4* in *Dfd*-depleted embryos (Fig EV5B). Together, these results showed that loss of specific homeodomain TFs, which are combinatorially expressed in single MNs, leads to cell-specific synaptic targeting defects and changes in target preferences of an individual MN at the end of embryogenesis.

The hypothesis that combinations of homeodomain TFs drive synaptic wiring and specificity implies that not only the loss of factors but also their ectopic activity should cause wiring defects. To test this assumption, we induced the expression of the homeodomain-containing *Hox* genes *lab*, *Dfd* and *Scr*, which are
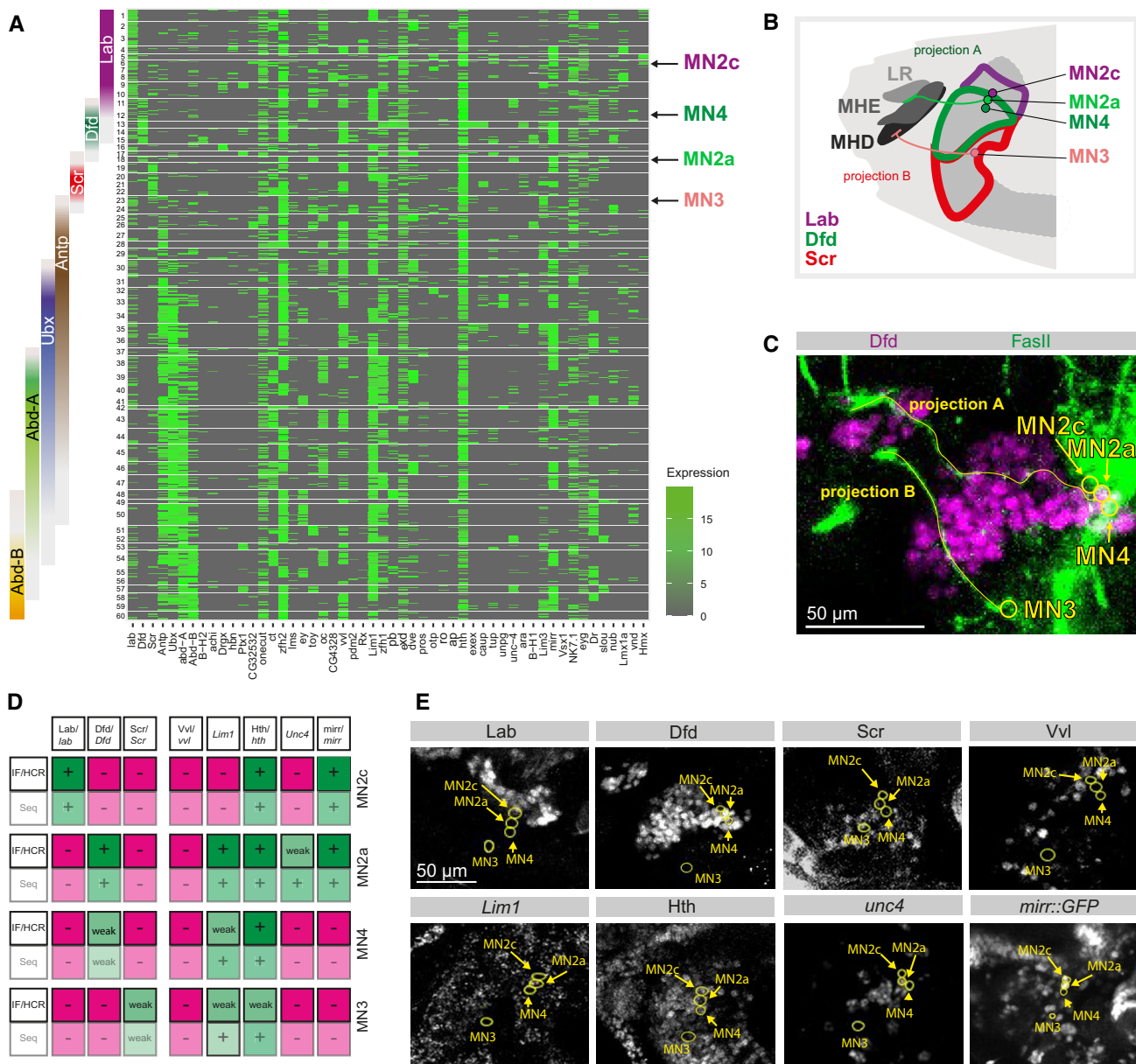
Figure 2. Specific combinations of homeodomain TFs are expressed in embryonic MNs during synaptic wiring.

A   Heatmap depicting the expression of homeodomain encoding genes (columns) across $n = 537$ single *twit*[low] MNs for which spatial mapping information was available (rows). Rows and columns are arranged by hierarchical clustering. Normalized expression levels are color-coded. Arrows indicate selected clusters for follow-up studies (see Fig 1D). *Hox* gene expression in individual cells as identified by scRNA-Seq is shown on the left by coloured bars, clusters 1 to 60 expressing specific combinations of homeodomain genes are shown on the left.

B   Schematic drawing of a stage 17 *Drosophila* embryonic head, highlighting the position of selected MNs (MN2c, MN2a, MN4, MN3) and feeding muscles (LR, MHE, MHD) located in consecutive segments expressing different Hox proteins (Lab, Dfd, Scr). Each MN innervates a specific muscle by a FasII-positive axonal projection, projection A connects MN2a and the MHE muscle, while projection B connects the MN3 with the MHD muscle.

C   Visualization of the different MNs based on Dfd expression in MN2a and the MHE, which is innervated by a FasII-positive axon emerging from MN2a (projection A), while the motoneuronal projections from MN3 (projection B) innervates the Dfd-negative MHD located just underneath the MHE (see also B). All MNs are highlighted by yellow circles.

D   Chart comparing the expression of eight homeodomain TFs in the four anterior MNs (MN2c, MN2a, MN4, MN3) as identified by scRNA-Seq (Seq) and immunohistochemistry (IF/HCR), with the absence of expression color-coded in magenta and the presence of expression in green. The genes were selected for validation, as they allowed an unambiguous distinction of the four anterior MNs according to the scRNA-Seq data.

E   Validation of the specific expression of the eight homeodomain TFs (Lab, Dfd, Scr, Vvl, Lim1, Hth, Unc4 and Mirr) in MN2c, MN2a, MN4 and MN3 by immunohistochemistry or *in situ* HCR. For expression analysis, proteins or transcripts were detected with specific antibodies (Lab, Dfd, Scr, Hth, Vvl) or HCR probes (Lim1, unc4) using wild-type embryos. Mirr expression was detected by GFP antibody stainings using a GFP fusion line (*mirr::GFP*). The stereotypic position of MN2c, MN4, MN2a and MN3 was identified by the two axon projections A and B (labelled by FasII), which invariantly innervate the MHE (projection A) and MHD (projection B) muscles in wild-type embryos. MN2c and MN4 are the two MNs adjacent to MN2a. The four MNs are highlighted by yellow circles.
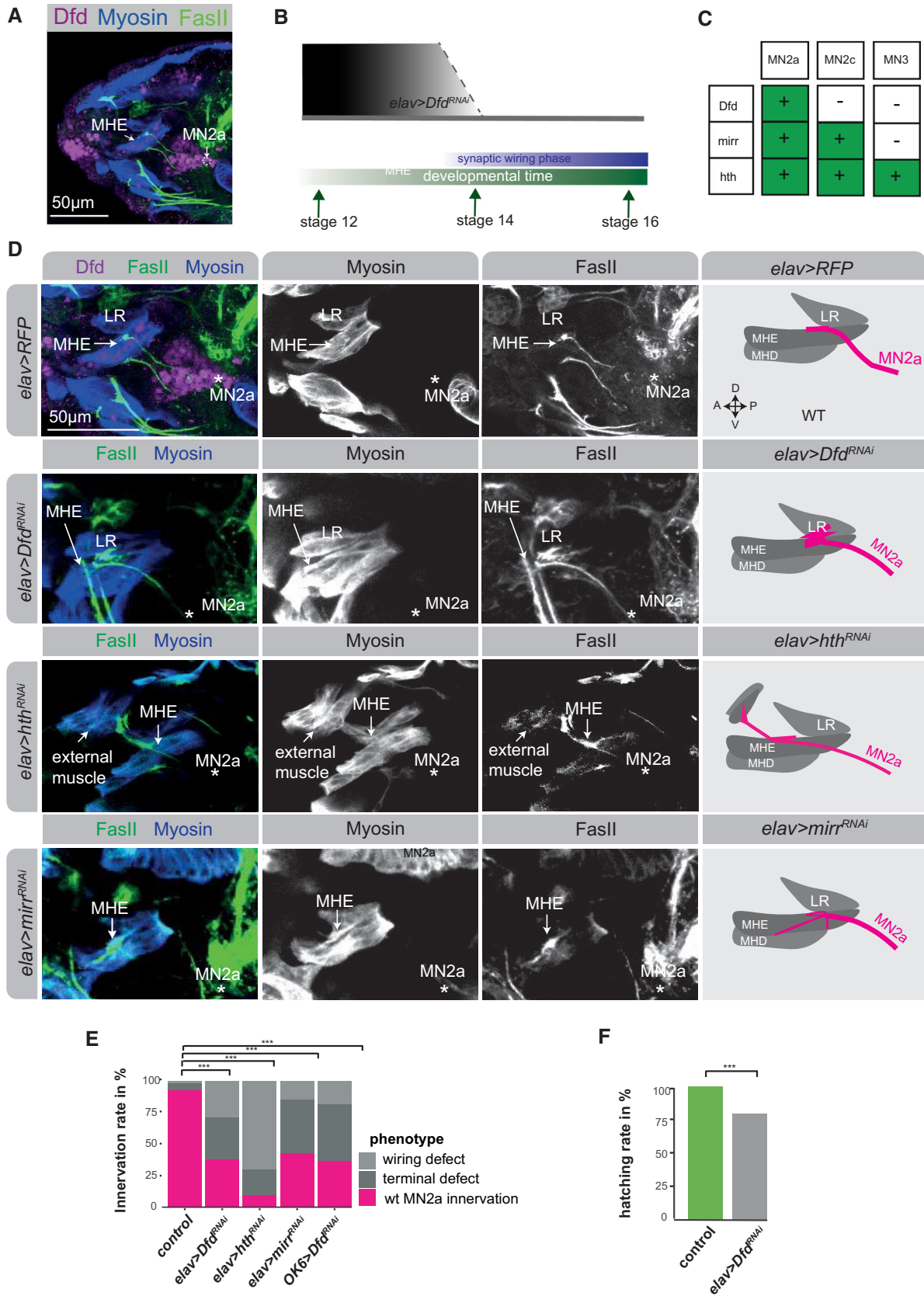
Figure 3.

Figure 3. Homeodomain TFs expressed in MNs fine tune wiring specificity in the neuromuscular system.

A Representative confocal image of a stage 17 embryonic head with Dfd-expressing cells (purple), Myosin expressing muscles (blue) and FasII-expressing axonal projections (green) highlighted. The MN2a is identified as a Dfd-expressing neuron projecting a FasII-expressing axon to the Dfd-positive MHE muscle. This stereotypic innervation is the basis for identifying wiring defects after RNAi interference shown in (D).

B Time course of RNAi induced effect using the pan-neural *elav-GAL4* driver line, showing that efficient knock-down of gene expression is achieved only at embryonic stage 14.

C Chart highlighting the expression of the three genes used for RNA interference, *Dfd*, *hth* and *mirr*, in MN2c, MN2a and MN3.

D Representative confocal images of stage 17 embryonic heads highlighting expression of Dfd (purple), Myosin in muscles (blue) and FasII in axonal projections (green) in control (*elav > RFP*) animals and after RNA interference with three different homeodomain TF genes, *Dfd*, *hth* and *mirr*, which are all co-expressed in MN2a (see Fig 2E). A zoom on the projection (FasII staining) of MN2a to the MHE and MHD muscle of an early stage 17 *Drosophila* embryo are shown. Asterisks highlight the location of MN2a, which is in all cases identified by the FasII expressing axonal projection emerging from a Dfd expressing MN, which normally innervates the Dfd-expressing MHE muscle (as shown in the *elav > RFP* control). The panel on the right side represents a schematic drawing of the confocal image shown on the left side, summarizing the innervation of the anterior muscles (LR, MHE, MHD) by projections emerging from MN2a (magenta) in control and perturbation conditions.

E Quantification of different phenotypes manifesting after RNA interference in comparison to control animals: MN2a axons projecting to the MHE muscle, representing the wild-type (wt) innervation pattern (magenta); MN2a axons projecting to muscles other than the MHE termed abnormal innervation (light grey); abnormal synaptic morphologies at MN2a-derived axon terminals termed terminal defects (dark grey). Note, each genetic experiment was performed in parallel to an adequate control experiment using the same driver line crossed to a line that controls expression of either UAS-RFP or UAS-GFP^{RNAi}. Each experiment was performed in triplicates, innervation rates were calculated from $n = 56$ for *elav > RFP*, $n = 21$ for *elav > Dfd*^{RNAi}, $n = 10$ for *elav > hth*^{RNAi} and $n = 7$ for *elav > mirr*^{RNAi}. In the case of Dfd knock-down, two different driver lines were used, the pan-neural *elav*-GAL4 and the motoneuron-specific *OK6*-GAL4 drivers, respectively. Both result in similar phenotypes, highlighting that the *elav*-GAL4-driven effects are specific to MNs. *P*-values between two genetic conditions were calculated by a two-sided Fisher test. ***$P < 0.005$.

F Correct MHE innervation is required for hatching of *Drosophila* embryos from the eggshell (Friedrich *et al*, 2016). Hatching rate was calculated based on the number of L1 larvae observed after 24 h in genetic crosses, depleted of Dfd (UAS-*Dfd*^{RNAi}) in neurons (*elav*-GAL4; $n = 217$) compared to crosses with control animals (mock = *elav > RFP*; $n = 156$). *P*-values were calculated by a two-sided Fisher test, ***$P < 0.005$.

active in consecutive segments in the *Drosophila* head (Fig 4A), in all neurons using the *elav*-GAL4 driver and investigated the innervation patterns of MN2a and MN3 (Fig 4A). In unperturbed conditions, the Dfd-expressing MN2a normally innervates the MHE muscle (Figs 2C and 3A), which also expresses Dfd (Fig 4A), while the Scr-expressing MN3 targets the Scr-positive MHD muscle (Figs 2 C and 4A). Ectopic expression of *Dfd* caused the MN3 to innervate the more anteriorly located Dfd-expressing MHE muscle, while ectopic *Scr* expression mis-guided the MN2a to the more posteriorly located Scr-positive MHD muscle (Figs 4B and C). Along the same lines, ectopic *lab* expression resulted in a high frequency of the MN2a to innervate the anteriorly located LR muscle (Fig 4B and C), which expresses Lab at low levels (Fig 4A). Intriguingly, this latter effect is reminiscent to the mistargeting phenotype observed in *Dfd* knockdown conditions (Fig 3D). These results are striking, as they show that the target preference of MNs can be changed along the AP axis in a predictive manner: mis-expression of more anterior *Hox* genes guided MNs to innervate more anterior muscles, while more posterior *Hox* genes directed the MNs to more posterior muscles (Fig 4B and C). However, these mistargeting effects were restricted, as motoneuronal axons hardly passed across muscles to innervate targets that are even further anteriorly or posteriorly located. For example, while ectopic *lab* expression allowed the MN2a to target the LR muscle (Fig 4B), the MN3 rarely innervated this muscle by crossing the Dfd-positive MHE muscle.

In sum, stage-specific genetic perturbation with homeodomain TFs resulted in cell-specific changes of muscle target selection by MNs, supporting the idea that these factors control the specificity of MN-muscle interactions in late embryonic stages when the first neuromuscular synaptic connections are established.

Functional relevance of Ig genes in mediating synaptic specificity

Our single-cell analyses and functional follow-ups revealed a critical role for combinations of homeodomain TFs in controlling synaptic

specificity in the neuromuscular system, motivating an investigation of potential downstream effectors. Unsupervised analysis of gene classes associated with the cell-specific homeodomain TF expression revealed that Ig encoding genes most strongly correlated with homeodomain clusters (Fig 5A and B). To visualize this relationship, we used each of the 60 homeodomain TF clusters identified before (Fig 2A) to calculate the corresponding expression of all Ig genes within each cluster (Fig 5C, Dataset EV1). Strikingly, this analysis revealed that each MN cluster expresses a complex but unique combination of Ig genes, with some Ig genes being expressed in all MNs, some in only a subset and a few exclusively in single MNs. For example, the Ig encoding gene *off-track (otk)* is expressed in posterior homeodomain clusters in RNA-Seq (Figs 5C and EV5D), which we confirmed by *in situ* HCR (Fig EV5C). Some other Ig encoding genes, such as those encoding Down syndrome cell adhesion molecules (Dscams) or Kekkon (Kek) neurotrophin receptors (Ulian-Benitez *et al*, 2017), are expressed in most homeodomain clusters, but are occasionally switched off in a few clusters. In other cases, for example genes of the DIP family, the Ig gene is specifically expressed in one or few homeodomain clusters (like *DIP-iota* in cluster 28; Fig 5A). This analysis also revealed that the four anterior MNs, MN2c, MN2a, MN4 and MN3, express specific combinations of Ig genes (Fig 5C).

To corroborate the regulation of Igs by homeodomain TFs, we manipulated the expression of the homeodomain encoding gene *Dfd* in neuronal cells and investigated the effects on Ig gene expression by qPCR or HCR. According to our scRNA-Seq data, *Dfd* is co-expressed with the Ig genes *DIP-gamma* and *DIP-kappa* in MN2a (Fig 5C, Dataset EV1). Both genes were downregulated when *Dfd* expression was reduced by RNAi in neural cells using the *elav*-GAL4 driver (Fig 5D and E). To provide further evidence that homeodomain TFs regulate Ig domain expression, we analysed previously generated whole-embryo Dfd ChIP-Seq data (Sorge *et al*, 2012) and Ubx ChIP-Seq data, which were retrieved from neuronal cells of late-stage embryos (Domsch *et al*, 2019). We found that Ig encoding
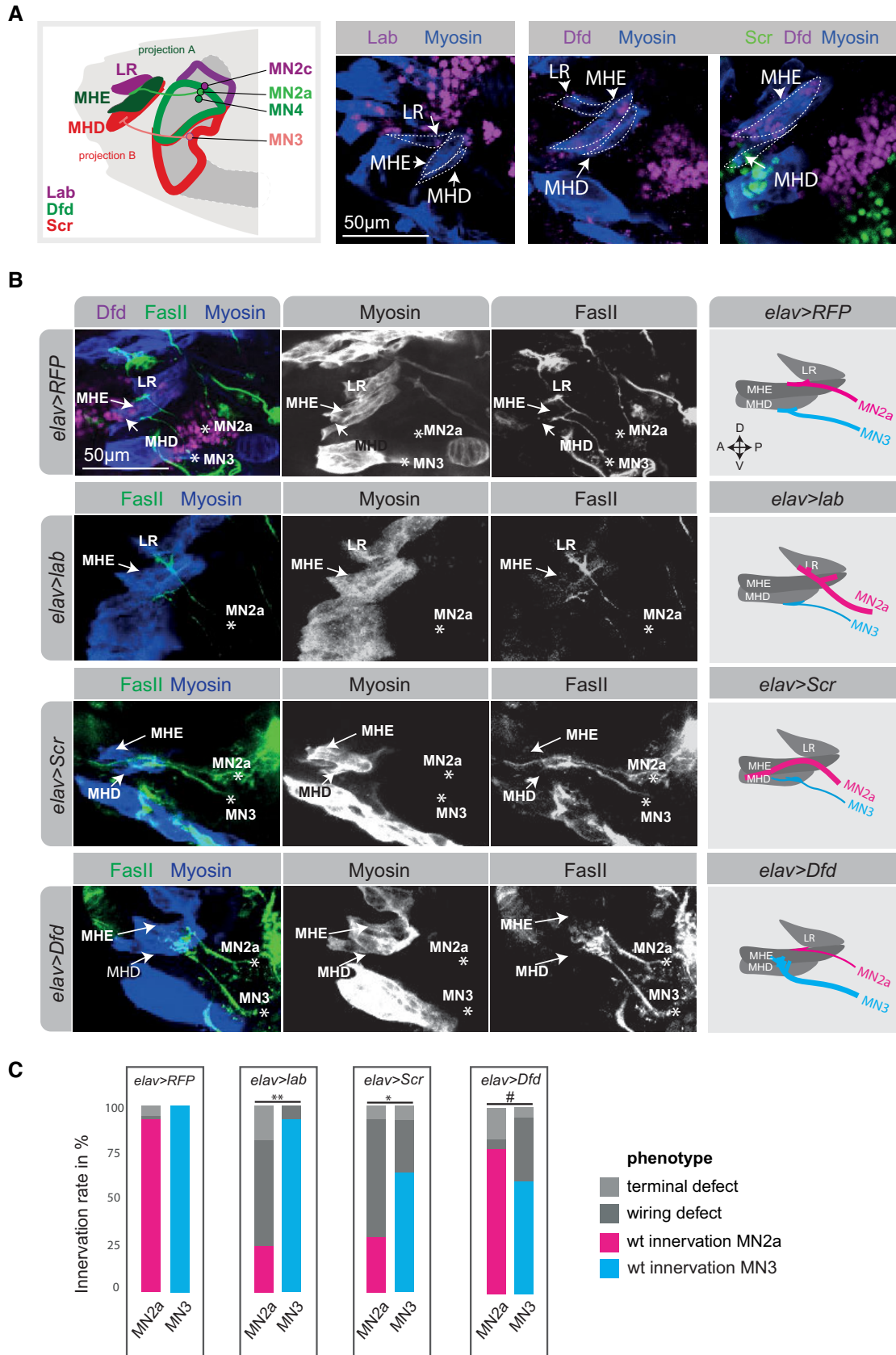
Figure 4.

◀

Figure 4. Ectopic homeodomain TF expression changes muscle target choice in a predictive manner.

A *Left panel*: Schematic drawing of a stage 17 *Drosophila* embryonic head, highlighting the stereotypic innervation of the anterior feeding muscles by specific MNs, MHE (green) by MN2a, which only expresses Dfd, MHD (red) by MN3, which only expresses Scr, and LR (purple) by MN2c, which only expresses Lab. *Right panel*: Representative confocal images of early stage 17 *Drosophila* embryonic heads, highlighting the expression pattern of the anterior Hox TFs Lab, Dfd and Scr, showing that Lab is expressed in the LR muscle, Dfd in the MHE muscle and Scr in the MHD muscle. The muscles are indicated by white dashed lines.

B Representative confocal images of stage 17 embryonic heads highlighting the expression of Dfd (purple), Myosin in muscles (blue) and FasII in axonal projections (green) in control (*elav > RFP*) animals and animals mis-expressing the Hox TFs Lab, Dfd and Scr by means of the *elav*-GAL4 driver. A zoom on the projections (FasII staining) of MN2a and MN3 to the MHE and MHD muscles of an early-stage 17 *Drosophila* embryo are shown. Asterisks highlight the location of MN2a and MN3, respectively. MN2a is identified by the FasII-expressing axonal projection emerging from a Dfd-expressing MN, which normally innervates the Dfd-expressing MHE muscle (as shown in the *elav > RFP* control), while MN3 is identified as the MN underneath MN2a, which normally innervates the MHD muscle (as shown in the *elav > RFP* control). The panel on the right side represents a schematic drawing of the confocal image shown on the left side, summarizing the innervation of the anterior muscles (LR, MHE, MHD) by projections emerging from MN2a (magenta) or MN3 (cyan) in control and mis-expression conditions.

C Quantification of different phenotypes manifesting in MN2a and MN3 after Hox mis-expression in comparison to control animals: MN2a axons projecting to the MHE muscle, representing the wild-type (wt) innervation pattern of MN2a (magenta); MN3 axons projecting to the MHD muscle, representing the wild-type (wt) innervation pattern of MN3 (cyan); MN2a or MN3 axons projecting to muscles other than the MHE or MHD termed wiring defects (dark grey); abnormal synaptic morphologies at MN2a- or MN3-derived axon terminals termed terminal defects (light grey). Note, each genetic experiment was performed in parallel to an adequate control experiment using the same driver line crossed to a line that controls expression of either UAS-*RFP* or UAS-*GFP*[RNAi]. Each experiment was performed in triplicates, innervation rates were calculated from $n = 16$ for *elav > lab*, $n = 18$ for *elav > Dfd* and $n = 25$ for *elav > Scr*. P-values between two genetic conditions were calculated by a two-sided Fisher test. [#]$P < 0.1$, *$P < 0.05$, **$P < 0.01$.

genes expressed in MNs overlap with the Dfd or Ubx bound targets (Fig 5F and G). In addition, homeodomain TF binding motifs were overrepresented within regulatory sequences associated with Ig encoding genes (Fig 5H), strongly supporting a direct regulation of Ig expression by homeodomain TFs (Fig 5I).

To investigate the functional role of Ig domain proteins in synaptic matching, we focused on the MN2a. This neuron co-expresses the homeodomain TFs Dfd and Mirr (but not Scr) and the Ig genes *DIP-kappa* and *DIP-gamma* (Fig 6B, Dataset EV1). RNAi-mediated knockdown of *DIP-kappa* and *DIP-gamma* in neural cells caused the MN2a to target the LR muscle (6% and 23%, respectively; Fig 6A and C) and additionally caused terminal defects (41% and 23%, respectively). Interestingly, this effect was similar to the phenotype induced by Dfd RNAi, but less pronounced (Fig 3D and E, 29% mistargeting and 33% terminal defects). This is consistent with the assumption that targeting specificity is mediated not by one but a combination of Ig genes acting downstream of homeodomain TFs. Importantly, ectopic expression of *DIP-gamma* in neural cells caused the MN3, which normally does not express this Ig gene, to innervate the *DIP-gamma* expressing MHE (Fig 6A and C), a phenotype we also observed when Dfd was ectopically expressed in MN3 by means of the neural *elav*-GAL4 driver (Fig 4B). Ectopic expression of other Ig genes that are not expressed in MN2a according to scRNA-Seq, such as *dpr1 or dpr11*, also showed MN-specific wiring defects towards alternative target muscles (Fig 6A and C).

Together, these results suggest that homeodomain TF combinations are associated with Ig gene expression, and Ig domain proteins are involved in mediating synaptic target specificity in the neuromuscular system downstream of homeodomain TFs and thereby modulate muscle target selection.

Homeodomain TFs mediate target specificity in interconnected MNs and muscle cells

The homeodomain TF Dfd is expressed in functionally connected cells of the feeding motor unit, the MN2a and the MHE muscle, while the adjacent LR and MHD muscles do not express Dfd (Fig 4A). Similarly, MN3 and the MHD muscle express the homeodomain TF Scr, whereas the more anterior located muscles and

neurons are devoid of Scr expression (Figs 2D and 4A). This observation led us to hypothesize that combinations of homeodomain TFs label matching synaptic partners at defined positions along the AP axis to control synaptic matching. To examine this hypothesis and complement the dataset on MNs, we performed scRNA-Seq of embryonic somatic muscles with enrichment for *Hox* genes as described above for MNs. To this end, we sorted somatic muscles using a fly stock expressing endogenously *GFP* tagged *Myosin heavy chain* (*Mhc-TAU-GFP*, (Chen & Olson, 2001). Analysis of this dataset by means of t-SNE and clustering indicates the existence of six relatively distinct subtypes of somatic muscle (Fig 7A). Using previously described marker genes, these were tentatively identified as dorsal somatic muscles (Dr positive), lateral somatic muscles (lms positive), ventral and lateral somatic muscles (mid and Poxm positive) and ventral somatic muscles (Ptx1 positive). For further analysis, we focused on the Poxm-positive cluster and demonstrated that homeodomain TFs and Ig domain encoding genes were highly variably expressed within this rather homogeneous somatic muscle subtype (Fig 7B). Most other clusters showed similar results (Fig 7C). Along the same line, we found binding of the homeodomain TF Ubx close to Ig genes expressed in mesodermal cells of late-stage embryos when analysing previously generated tissue-specific Ubx ChIP dataset (Domsch *et al*, 2019) (Fig 7D). These findings strongly suggest that homeodomain TF combinations are associated with Ig gene expression in MNs as well as in muscles.

To support the hypothesis that co-expression of identical homeodomain TFs in interconnected cells is critical for synaptic matching, we interfered with *Dfd* in neurons (*elav*-GAL4) or muscles (*Mef2*-GAL4) or in both tissues (*elav*-GAL4; *Mef2*-GAL4;) and analysed the innervation of anterior muscles (LR, MHE and MHD) by anterior MNs. In all three conditions, we observed significant synaptic wiring defects and terminal defects with a significant increase in phenotypes when *Dfd* was knocked down in both tissues. While neural *Dfd* depletion led to mistargeting of the MN2a to the LR muscle (Fig 8A and B), *Dfd* knock-down in muscles changed the innervation of the MHE muscle, which normally expresses Dfd, as it was now targeted by MNs that normally innervated the MHD muscles (Fig 8A and B). Knockdown of *Dfd* in both muscle and neurons
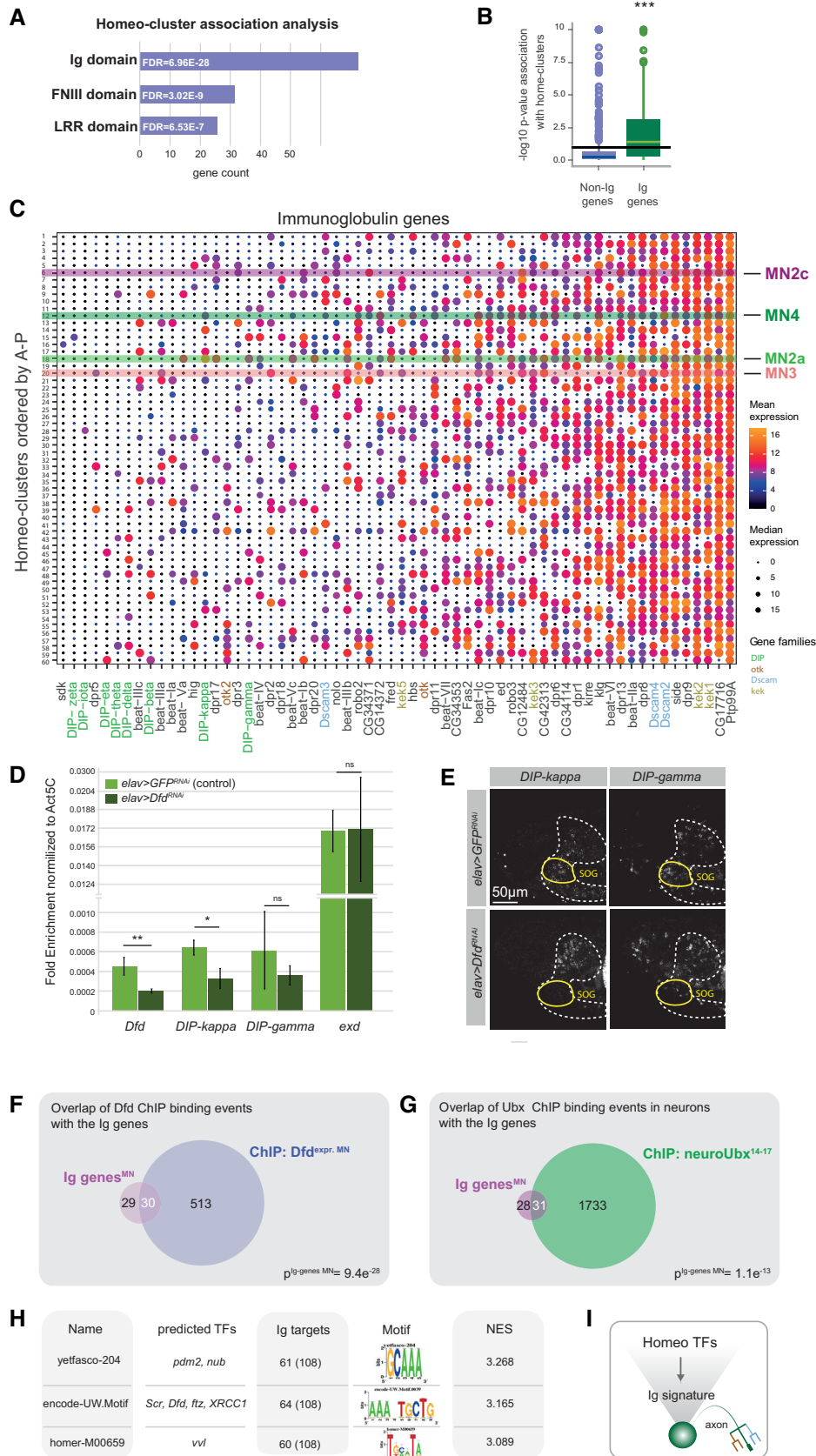
Figure 5.

Figure 5. Homeodomain TFs regulate Ig genes expressed in embryonic MNs.

A, B   Analysis of gene classes associated with the homeodomain TF clusters. For each gene in the dataset, normalized, scaled expression in single cells was modelled as a function of homeodomain TF cluster identity, and significance of the association was determined using an F-test. Panel (A) depicts the number of genes with significant homeodomain TF cluster association falling into distinct gene classes. P-values were retrieved by a Fisher test. Panel (B) contrasts the P-values for homeodomain TF cluster association between Ig domain genes and other genes using a boxplot. See Materials and Methods, section *Data visualization* for a definition of boxplot elements. P values are from a two-sided Wilcoxon test. ***P < 0.005.

C   Single cells were grouped into 60 clusters according to their expression of genes encoding homeodomain proteins (Fig 2A) and arranged along the inferred AP position (Fig 1D, 1E). For each cluster, the mean expression of Ig encoding genes was computed. Heatmap depicts Ig encoding genes in columns and clusters in rows; mean expression is color-coded and median expression level is visualized by circle size. Gene families are highlighted in colour codes. The coloured bars highlight the Ig genes expressed in anterior MNs (MN2c, MN2a, MN4, MN3).

D   Quantification of Ig gene expression changes after interference with *Dfd* expression in neural cells (*elav* > *Dfd^RNAi^*) by qPCR. Two genes, *DIPgamma* and *DIPkappa*, were selected as potential Dfd targets based on their co-expression in Dfd-positive MNs. The bar chart shows the fold enrichment after normalization against Act5C of the *Dfd* and *Ig* genes in control versus *elav* > *Dfd^RNAi^* conditions from three biological replicates and shows that *Dfd* and *DIP-kappa* expression are significantly reduced in *elav* > *Dfd^RNAi^* embryos, whereas the expression of the homeodomain TF gene *exd* is unchanged (internal control). P-values were calculated from a t-test: *<0.05, **<0.01, ns non-significant. Error bar denotes standard deviation.

E   HCR for *DIPgamma* and *DIPkappa* transcripts in *Drosophila* stage 17 embryos. RNA expression of both genes is analysed in control (*elav* > *GFP^RNAi^*) versus Dfd-depleted (*elav* > *Dfd^RNAi^*) animals, the ventral nerve cord is highlighted by a white dashed line, the Dfd-expressing suboesophageal ganglion (SOG) is indicated by a yellow circle.

F   Venn diagram displaying an overlap of Ig domain genes expressed in MNs and genes that are bound by Dfd based on the Dfd ChIP-Seq performed in a whole embryo (Sorge *et al*, 2012). P-value was calculated using a hypergeometric test using as a reference all protein coding genes *n* = 13,920.

G   Venn diagram displaying the overlap of Ig domain genes expressed in MNs that are bound by Ubx identified in a neuronal tissue-specific ChIP-Seq (Domsch *et al*, 2019). P-value was calculated as described in (F).

H   iRegulon analysis was used to identify TF motifs enriched in the vicinity of Ig encoding genes. Three of the top 15 highest ranked motifs of TFs are shown; the predicted targets of these motifs are homeodomain TFs.

I   Schematic drawing of the relationship between the homeodomain and the Ig molecule clusters in a single neuron.

caused aberrant innervations of the MHE muscles and loss of the stereotypic innervation (Fig 8A and B).

In sum, the results are in line with the hypothesis that combinations of homeodomain TFs are required to align synaptic partners in the neuronal and muscle tissues and coordinate their preferences in synaptic partner choice.

Discussion

Each neuron in the nervous system chooses a single target cell from a high number of possible interactions, which is a prerequisite for the formation of stereotypic neuronal circuits. This extraordinary degree of precision is thought to be mediated by a matching set of cell specifically expressed cell recognition and adhesion molecules such as CSPs (Sanes & Zipursky, 2020). However, up to date, a systematic scrutinization of such a "connectivity map" and the upstream mechanisms fine-tuning its expression are missing. The most important reasons for this are the lack of single cell-specific neuronal markers, the complexity of neuronal systems and the possibly gradual and combinatorial nature of CSP expression and function.

Our experimental design allowed us to overcome these challenges. First, by focusing on one neuronal subtype, the motoneuronal population in *Drosophila* embryos, we were able to reduce neuronal complexity. Second, we investigated and manipulated this cell population exactly at the time when they form the first stereotypic connections with their muscle targets during embryogenesis without affecting earlier stages of development. This approach allowed us to identify molecular cues critical for synaptic wiring, as molecules driving this process are known to be most differentially expressed at that time (Li *et al*, 2017). Third, MNs form highly cell-specific connections with muscles and are present in a relatively small number per embryo. Thus, we used a single-cell genomic

approach with a high number of biological replicates of every biologically unique cell. Thereby, we identified novel markers specific to single cells or small groups of cells that in turn permit the identification of transcriptome signatures relevant for wiring. And finally, we implemented a spatial mapping approach based on *Hox* gene expression to locate MNs along their AP position and thereby gained insight into the role of spatial mechanisms during synaptic wiring.

Our scRNA-Seq data revealed that a cell-specific combination of homeodomain TFs acts as a major component of transcriptional heterogeneity during the wiring phase of MNs. Despite the use of a high number of biological replicates for each unique cell, scRNA-Seq on its own cannot be applied to unanimously identify biologically unique cells. Thus, we used imaging to demonstrate in four cases a correspondence between homeo-clusters identified by scRNA-Seq and gene expression patterns in stereotypically defined single cells. Additionally, we showed that ectopic expression of homeodomain TFs caused wiring defects in MNs not expressing the homeodomain TF, whereas knockdown caused wiring defects in MNs expressing the homeodomain TF. Together, these data allowed us to conclude that cellular heterogeneity within postmitotic *Drosophila* MNs could be described by cell-specific combinations of homeodomain TFs, which we showed to control late events in neuronal differentiation, in particular synaptic wiring, at the single-cell level. Thus, our unbiased genomic approach suggested that homeodomain TFs could be major drivers in the synaptic wiring process, which is in line with many previous studies showing that this TF class plays an important role in MNs in different organisms (Thor *et al*, 1999; Jurata *et al*, 2000; Thor & Thomas, 2002; Sanguinetto *et al*, 2008; Dasen & Jessell, 2009; Arber, 2012; Philippidou *et al*, 2012; Zarin *et al*, 2014; Meng & Heckscher, 2020; Hobert, 2021).

Homeodomain TF expression is not only highly specific to individual MNs, but is also associated with profound differences in the entire transcriptome and in particular the expression of Ig CSPs as
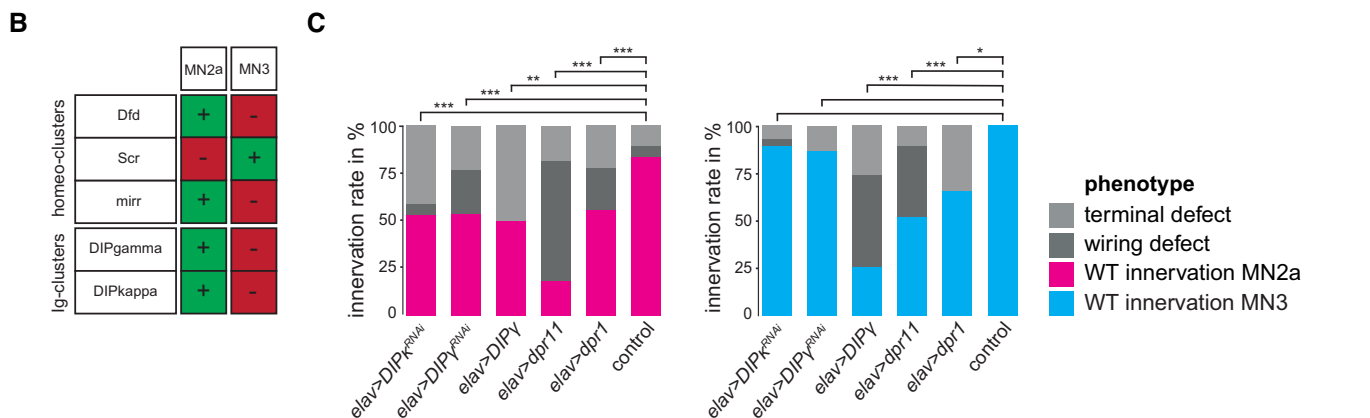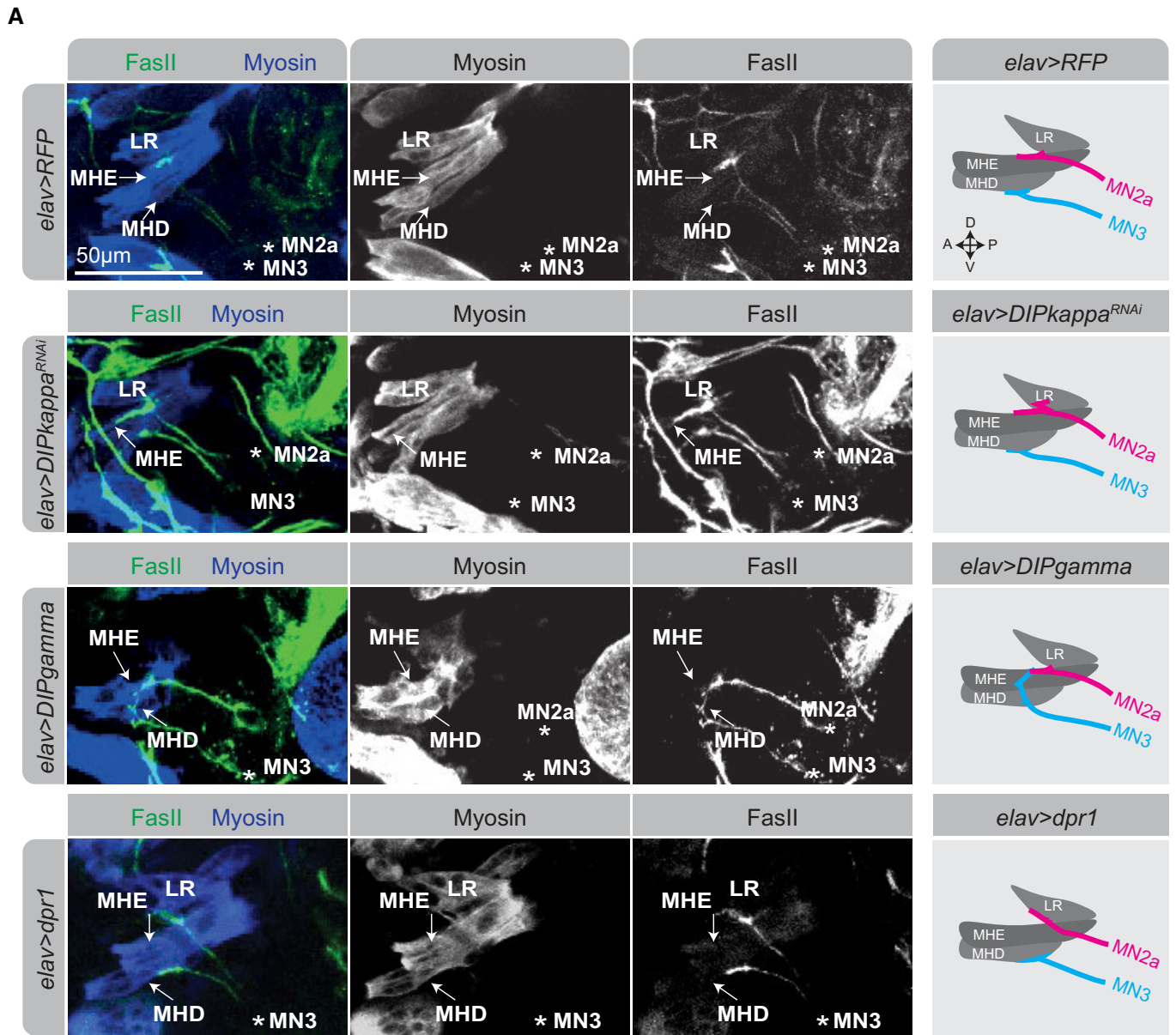
A

B

C

Figure 6.

◀

possible effectors of synaptic specificity. Thus, we used homeodomain TF expression as classifiers to define cellular identities and systematically investigate the distribution of CSPs between cells. We found on average 5–10 homeodomain TF genes and 20–40 Ig genes to be cell specifically expressed in individual MNs, revealing that Ig CSP expression is highly combinatorial and even more complex than the expression of the corresponding homeodomain TFs. Although reported previously that synaptically connected neurons express multiple ligand–receptor pairs (Tan *et al*, 2015), our data suggest that this process is maybe even more complex than previously thought. Interestingly, most CSPs change more gradually between cells, and unlike in the case of homeodomain TF expression, binary expression patterns of Ig genes were not observed in the data (Wit & Ghosh, 2016; Oostrum *et al*, 2020). In line, previous studies on single molecules already indicated that small changes in relative expression levels of CSPs in matching partners can change synaptic specificity (Sweeney *et al*, 2011; Yogev & Shen, 2014). In addition, we found in our dataset, some Ig genes, such as *Dscam* encoding genes, to be broadly expressed but not present in some defined cells, while many other Ig genes are less specific for cellular identities. Importantly, only one Ig molecule class, the DIP genes, were found to be distinct for specific cellular identities. Intriguingly, previous expression and functional studies had already implicated these molecules as strong candidates for synaptic targeting function (Carrillo *et al*, 2015; Tan *et al*, 2015; Ashley *et al*, 2019; Cheng *et al*, 2019), and we now provide additional evidence that they are involved in controlling synaptic matching in the embryonic neuromuscular system. Taken together, our data strongly support a model whereby a highly combinatorial Ig gene expression program active in single cells drives synaptic specificity and connectivity between cells. This complexity might explain why null mutants in many genes contributing to specificity have been shown to result in low penetrance or incomplete phenotypes (Xu *et al*, 2018, 2019). Many recognition molecules might function redundantly in the synaptic matching process either by acting in parallel pathways or in protein complexes. We envision such a complexity to ensure a higher robustness in synaptic matching even when single components are missing or expressed at reduced levels. In the future, it will be important to study the contribution of CSP combinations to synaptic

wiring at the single-cell level to resolve redundancy and specificity within this system.

How is the complex CSP expression program critical for synaptic wiring controlled in individual MNs? Several lines of evidence indicate that combinations of homeodomain TFs are the direct upstream regulators of CSPs in individual MNs. It has been shown before for some homeodomain TFs that they control the precise matching of MNs and their target muscles (Jurata *et al*, 2000; Thor & Thomas, 2002; Sanguinetto *et al*, 2008; Dasen & Jessell, 2009; Arber, 2012; Philippidou *et al*, 2012; Zarin *et al*, 2014; Meng & Heckscher, 2020; Hobert, 2021). In addition, some previous data provided already evidence for a functional link between the expression of receptor molecules and homeodomain TFs (Labrador *et al*, 2005). We now show that manipulation of homeodomain TF expression during the wiring phase by knockdown or ectopic expression leads to changes in target preferences in line with the expression pattern of the respective TF. For example, mis-expression of more anterior *Hox* genes guided MNs to innervate an anterior muscle, while mis-expression of posterior *Hox* genes guided MNs to innervate a posterior muscle. Importantly, phenotypes induced by knockdown or overexpression of one Hox TF are phenocopied by manipulations of its putative Ig CSP targets. Consistently, we found the expression of Ig CSP encoding genes to be changed in the absence of homeodomain TFs. Furthermore, chromatin immunoprecipitation studies indicate a direct interaction of homeodomain TFs with control regions of their Ig targets. Together, the combination of single-cell genomics (co-expression), chromatin immunoprecipitation (binding of enhancers) and genetics (phenocopies, changes in expression of downstream genes) strongly suggest that cell-specific homeodomain TF programs are important regulators of synaptic specificity and that combinations of Ig proteins act directly downstream of homeodomain TFs to mediate their function. How these complex combinations of Ig molecules are controlled in a cell-specific manner by homeodomain TFs will be an important problem to solve in the future. This is particularly puzzling for a TF class like the homeodomain TFs, which interact with highly similar DNA sequences (Ekker *et al*, 1994; Noyes *et al*, 2008). Whether other TF classes, although not identified as highly variable genes but expressed in our MN transcriptomes, contribute to specificity in Ig gene expression will be one of the many paths to follow
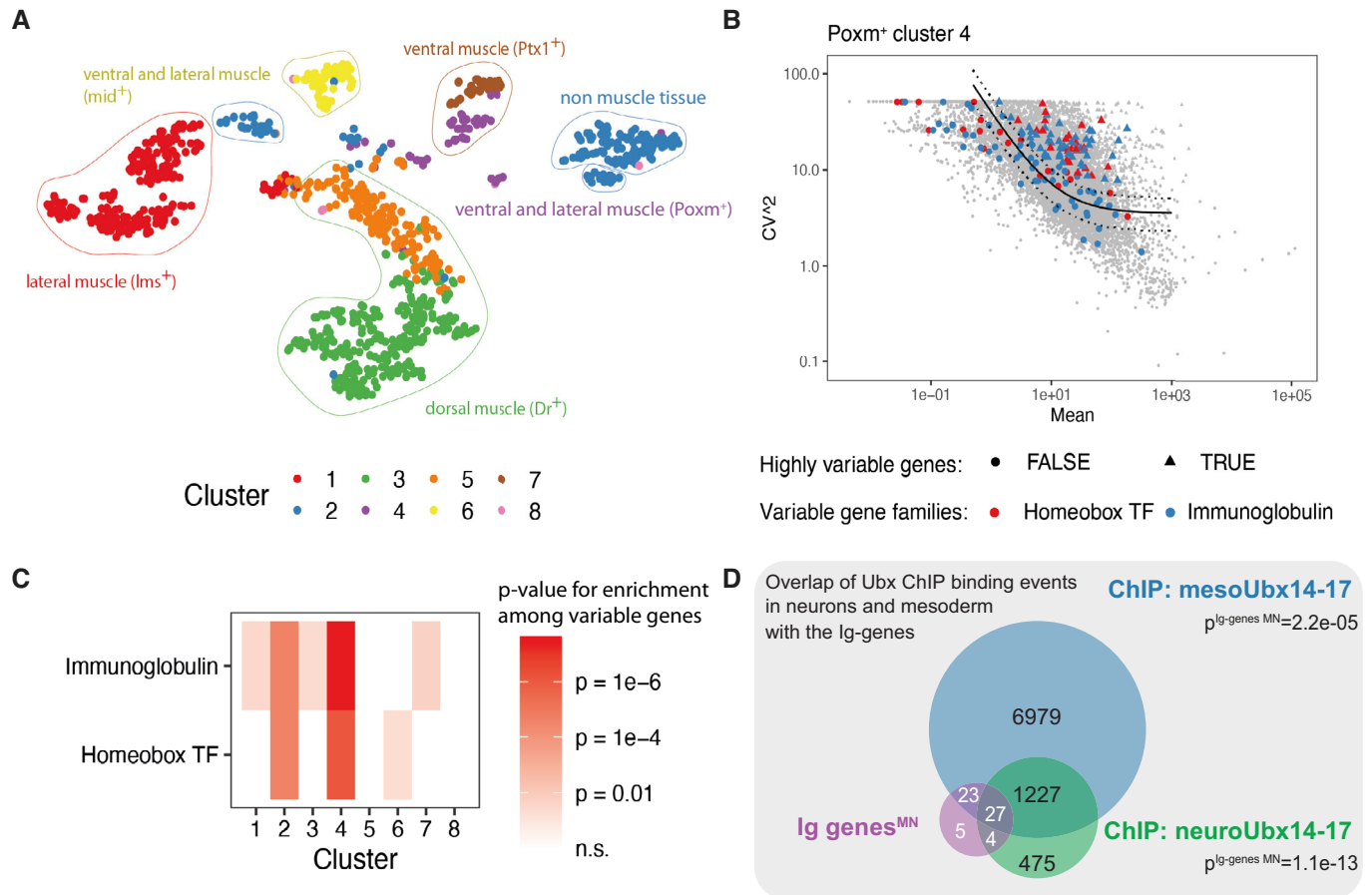
Figure 7. Homedomain TFs are differential expressed in embryonic muscles.

A t-distributed stochastic neighbour embedding (t-SNE) plot of single-cell transcriptomes of GFP expressing somatic muscle cells sorted from *Mhc-TAU-GFP* expressing stage 16 *Drosophila* embryos. Colours correspond to clusters identified using hierarchical clustering that was annotated using marker gene expression of muscle subtypes.

B Identification of highly variable genes using the method by (Brennecke *et al*, 2013). Scatter plot depicts for each gene the mean expression and squared coefficient of variation across cells using the Poxm⁺ cluster as input. The solid line indicates the fit, dashed lines the 95% confidence interval. Genes with a significantly elevated variance are shown as triangles, other genes as circles. Different gene classes are colour-coded.

C Statistical model of Brennecke *et al* (2013) was used to identify highly variable genes. Colour code displays enrichment of Ig and homeodomain TF-encoding genes among the variable genes according to a hypergeometric test.

D Venn diagram displaying an overlap of Ig domain genes bound by Ubx in muscle and neuronal tissues, identified in a tissue-specific ChIP-Seq (Domsch *et al*, 2019). *P*-value was calculated using a hypergeometric test using as a reference all protein coding genes $n = 13,920$.

based on the resources generated in this study. Our analysis also revealed that homeodomain TFs, although co-expressed in the same MN, can have different functions in synaptic wiring, as their knockdown does not necessarily result in the same wiring defects (Fig 3). This furthermore highlights the combinatorial nature of regulating synaptic specificity. In future, it will be highly relevant to resolve this combinatorial control on a single-cell level. Finally, we would predict that synaptic wiring defects resulting from the loss of one homeodomain TF(s) cannot be rescued by the expression of a single Ig target but very likely requires changes in the expression of multiple of these molecules.

In sum, we propose a model whereby the position of every cell is imprinted early in embryonic development by patterns of homeodomain TFs (Fig 9). Expression patterns of these factors become more complex and combinatorial with each cell division until small groups of cells and possibly every single cell are uniquely defined by a homeodomain TF program. These unique combinations of homeodomain TFs then in turn regulate specific downstream programs of Ig gene expression. In the target cells (here: muscles), a similar TF combination is an important determinant of connectivity and possibly regulates the expression of a complementary Ig receptor expression program. This molecular logic enables every single cell to find its corresponding interaction partner based on complementary adhesive properties mediated by combinations of Ig domain molecules. Thus, this concept explains how a molecular memory of cell body position is translated into invariant cell–cell adhesion events by means of a linked homeodomain TF–Ig program. In the future, the combinatorial manipulation of CSPs and homeodomain TFs in single cells (Replogle *et al*, 2020) will allow a further functional dissection of their interactions and regulatory mechanisms.
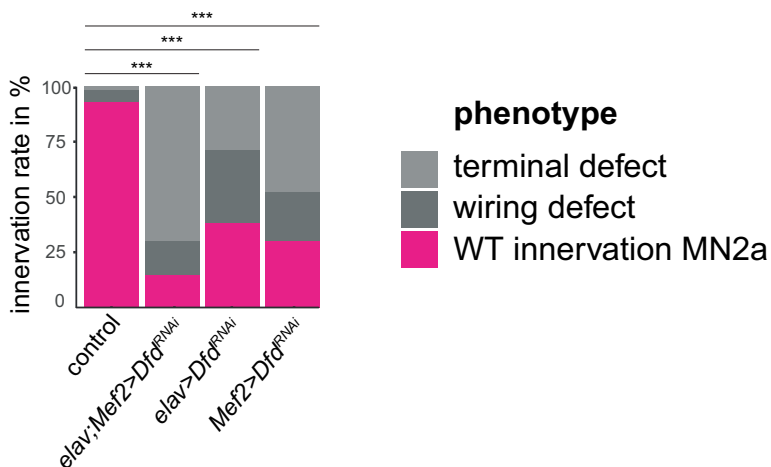
A
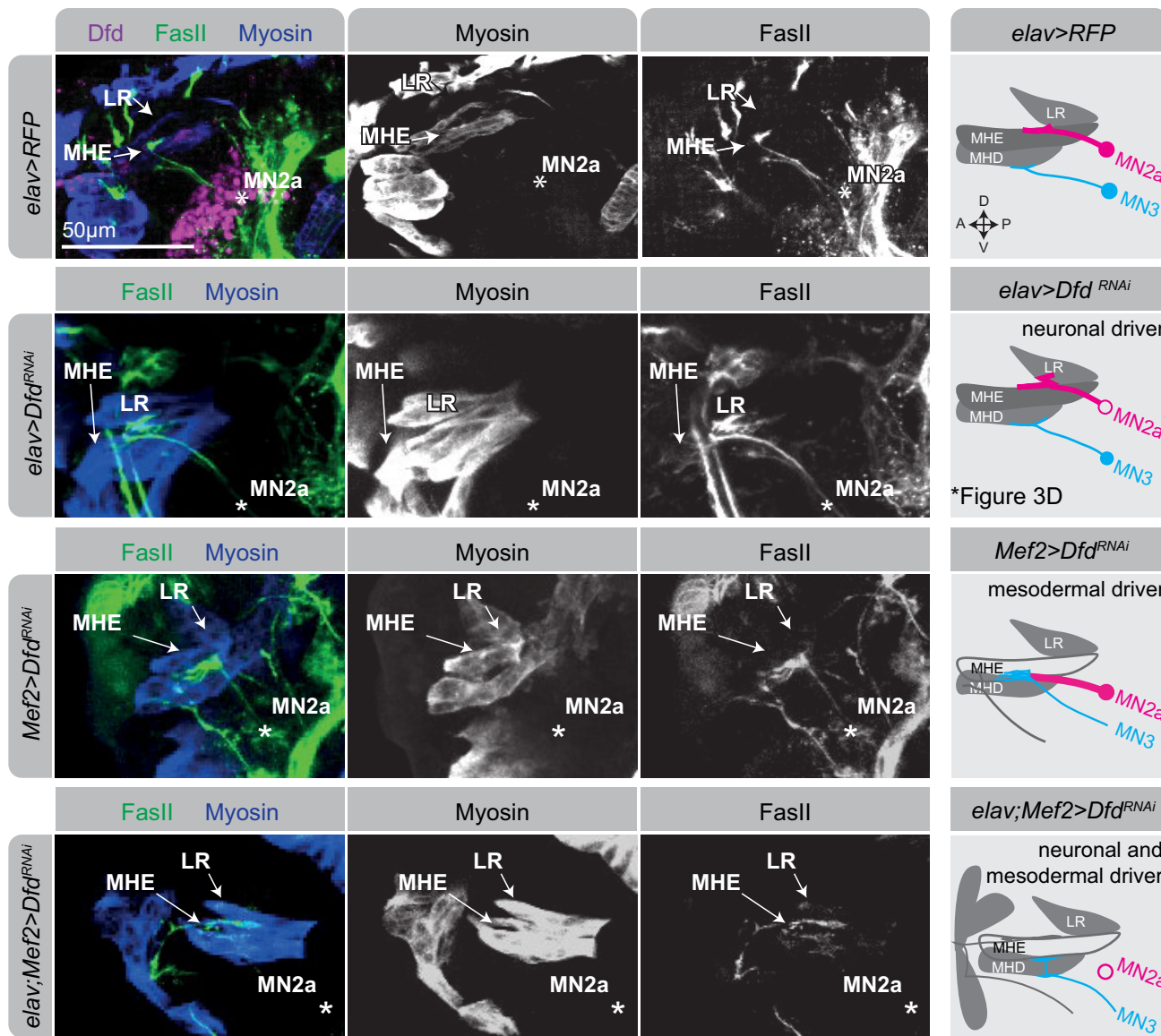
B

Figure 8.

Figure 8. Synaptic targeting requires the homeodomain TF Dfd in neurons and muscles.

A   Representative confocal images of stage 17 embryonic heads highlighting the expression of Dfd (purple), Myosin in muscles (blue) and FasII in axonal projections (green) in control (*elav > RFP*) animals and in animals with reduced *Dfd* expression in neurons (*elav > Dfd^RNAi*) or in muscles (*Mef2 > Dfd^RNAi*) or both in both tissues (*elav;Mef2 > Dfd^RNAi*). The single channels for the projections (FasII staining) of MN2a and/or MN3 to the MHE, MHD or LR muscles of an early-stage 17 *Drosophila* embryo are shown. Asterisks highlight the location of MN2a and MN3, respectively. MN2a is identified by the FasII-expressing axonal projection emerging from a Dfd-expressing MN, which normally innervates the Dfd-expressing MHE muscle (as shown in the *elav > RFP* control), while MN3 is identified as the MN underneath MN2a, which normally innervates the MHD muscle (as shown in the *elav > RFP* control). The panel on the right side represents a schematic drawing of the confocal image shown on the left side, summarizing the innervation of the anterior muscles (LR, MHE, MHD) by projections emerging from MN2a (magenta) or MN3 (cyan) in control and perturbation conditions.

B   Quantitative analysis of phenotypes. Each experiment was performed in triplicates, innervation rates were calculated from $n = 21$ for *elav > Dfd^RNAi*, $n = 23$ for *Mef2 > Dfd^RNAi* $n = 20$ for *elav;Mef2 > Dfd^RNAi* animals. Note, each genetic experiment was performed in parallel to an adequate control experiment using the same driver line crossed to a line that controls expression of either *UAS-RFP* or *UAS-GFP^RNAi*, P-values between two genetic conditions were calculated by a two-sided Fisher test. ***$P < 0.005$.
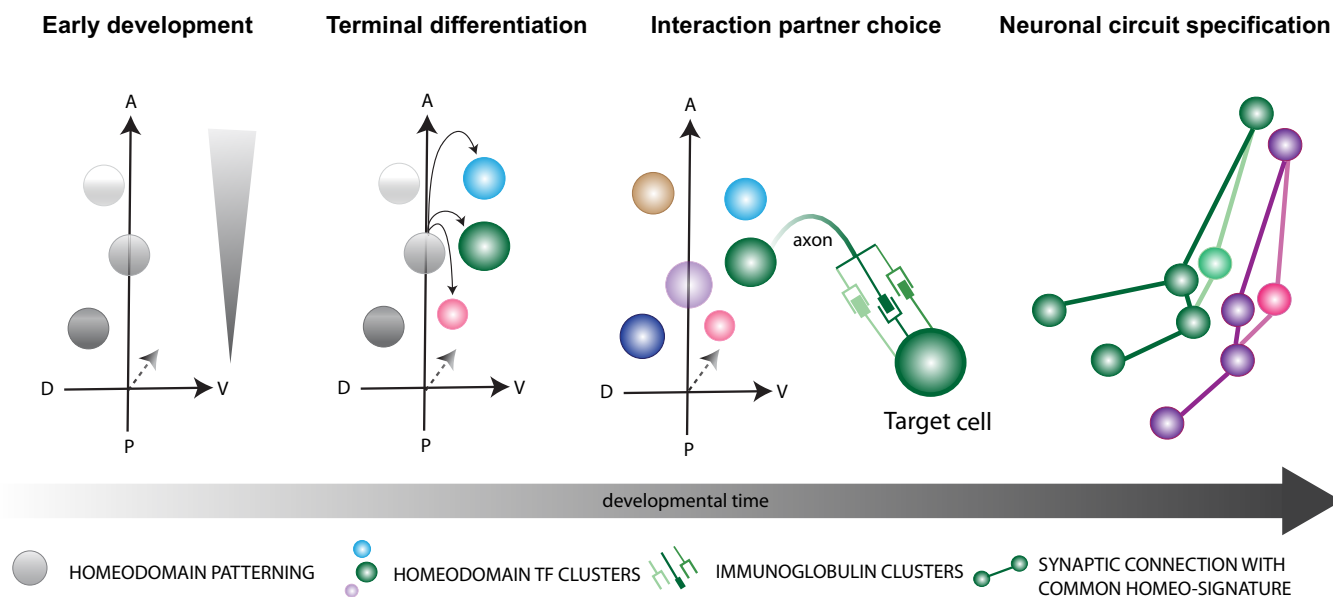

Figure 9. Model depicting the molecular mechanism for stereotypic synaptic partner selection.

Left panel: Early developmental programs pattern single cells according to their position in the embryo. Morphogenic gradients establish specific homeodomain protein expression patterns along embryonic body axes (i.e. dorsal ventral axis = DV axis, anteroposterior axis = AP axis). Left middle panel: After neuroblasts have undergone several divisions, neurons terminally differentiate and establish unique identities depending on their homeodomain TF expression (colour code). Right middle panel: A unique homeodomain TF expression specifies Ig domain receptor expression in matching synaptic partners, resulting in differential affinities of Ig domain proteins promoting selective synaptic target choice. Right panel: Common homeo-signatures regulate interaction of partners within individual neuro-muscular circuits.

Materials and Methods

Reagents and Tools table

Reagent/Resource	Reference or source	Identifier or Catalog Number
Experimental Model: *Drosophila melanogaster*		
UAS-EGFP RNAi (y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8] = VALIUM20- EGFP.shRNA.3}attP2)	Bloonington Drosophila Stock Center	41560
UAS-mirr RNAi (y[1] sc[*] v[1]; P{TRiP.HMC06139}attP2)	Bloonington Drosophila Stock Center	65877
UAS-hth RNAi (P{y[1] sc[*] v[1]sev21; TRiP.HMS01112}attP2)	Bloonington Drosophila Stock Center	34637

Reagents and Tools table   (continued)

Reagent/Resource	Reference or source	Identifier or Catalog Number
UAS-DIP gamma RNAi (y[1] sc[*] v[1]; P{TRiP.HMS06062} attP40)	Bloonington Drosophila Stock Center	80461
UAS-DIP kappa RNAi (y[1] sc[*] v[1]; P{TRiP.HM04050}attP2)	Bloonington Drosophila Stock Center	31740
UAS-Dfd RNAi (UAS-Dfd-si on chr.III)	Friedrich e a (2016)	N/A
UAS-mCD8::GFP (y[1] w[*]; P{w[+mC] = UASmCD8::GFP.L}LL5, P{UAS- mCD8::GFP.L}2)	Bloonington Drosophila Stock Center	5137
UAS-mRFP (w[1118]; P{w[+mC]=UAS-myr-mRFP}1)	Bloonington Drosophila Stock Center	7118
UAS-lab (w[1118]; P{w[+mC]=UAS-lab.M}X2)	Bloonington Drosophila Stock Center	7300
UAS-dpr1 (y[1] w[*]; P{UAS-dpr1.N}2/CyO)	Bloonington Drosophila Stock Center	25081
UAS-Scr (w[1118]; P{w[+mC]=UAS-Scr.M}EE2/TM6B, Tb[+])	Bloonington Drosophila Stock Center	7302
M{UAS-DIP-γ.ORF.3xHA.GW}ZH-86Fb	Flyorf	F003086
M{UAS-dpr11.ORF.3xHA.GW}ZH-86Fb	Flyorf	F003032
UAS-Dfd-HA	Katrin Domsch (designed for this study)	N/A
elav-Gal4 (P{w[+mW.hs]=GawB}elav[C155])	Bloonington Drosophila Stock Center	458
Mef2-Gal4	Ranganayakulu et al (1998)	N/A
elav-Gal4; Mef2-Gal4	designed for this study	N/A
Ok6-Gal4	Aberle et al (2002)	N/A
Ok371-Gal4	Mahr and Aberle (2006)	N/A
lab::GFP (y[1] w[*]; P{y[+t7.7] w[+mC]=lab-GFP.FPTB}attP40)	Bloonington Drosophila Stock Center	66551
Frq1MI00709-GFSTF (y[1]w[*]; Mi{PT-GFSTF.0}Frq1MI00709-GFSTF.0)	Bloonington Drosophila Stock Center	60284
mirr-GFP (y[1]w[*]; P{mirr-GFP.FPTB}attP40)	Bloonington Drosophila Stock Center	68183
MHC-TAU-GFP	Chen and Olson (2001)	N/A
Antibodies		
Rabbit anti-Hth	Richard Mann lab (McKay et al, 2009)	N/A
Rabbit anti-GFP	invitrogen	A11122
Rabbit anti-Lab	Thomas Kaufmann lab (Diederich et al, 1989)	N/A
Rat anti-Myosin	abcam	ab51098-100
Rat anti-GFP	chromotek	3H9
Rat anti-Vvl	W.A Johnson lab (Anderson et al, 1995)	N/A
Rat anti-ELAV	Developmental Studies Hybridoma Bank - DSHB	7E8A10
Goat anti-Abd A	Santa Cruz	27063
Guinea pig anti-Dfd	Katrin Domsch (designed for this study)	N/A
Guinea pig anti-Ubx	Katrin Domsch (Domsch et al, 2019)	N/A
Guinea pig anti-Vnd	Robert Zinzen	N/A

Reagents and Tools table   (continued)

Reagent/Resource	Reference or source	Identifier or Catalog Number
Mouse anti-Scr	Developmental Studies Hybridoma Bank - DSHB	6H4.1
Mouse anti-Abd A (all isoforms)	Developmental Studies Hybridoma Bank - DSHB	Ubx/ABD-A FP6.87
Mouse anti-And B	Developmental Studies Hybridoma Bank - DSHB	1A2E9
Mouse anti-Antp	Developmental Studies Hybridoma Bank - DSHB	8c11
Mouse anti-FasII	Developmental Studies Hybridoma Bank - DSHB	1D4
Oligonucleotides and sequence-based reagents		
oligo-dt30VN	This study	AAGCAGTGGTATCAACGCAGAGTACT30VN
TSO	This study	AAG CAG TGG TAT CAA CGC AGA GTA CAT rGrG+G
IS PCR primer	This study	AAG CAG TGG TAT CAA CGC AGA GT
targeted_Dfd_RT_rv	This study	TCG GAT TGT TGC TGT TGA AG
targeted_Ubx_RT_rv	This study	CAG AAT TTT GCT CGC ATT CA
targeted_AbdA_RT_rv	This study	CAT GCG TTG CTC TAT CAA A
targeted_AbdB_RT_rv	This study	AAT ATA ATG CTC GGG GCA AA
targeted_Scr_RT_rv	This study	ATT GGG CGA TAC AAA CGA AG
targeted_Lab_RT_rv	This study	CCC TTC AAC TTT GCT TGC TC
targeted_Antp_RT_rv	This study	AAC CAT ACC CAG TCC ACC AA
targeted_Dfd_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTC CCT GGA TGA AGA AGA TCC A
targeted_Ubx_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTA AGG AGC TGA ACG AAC AGG A
targeted_AbdA_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTC TGG ACA AGA GCA ATC ACG A
targeted_AbdB_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTC GGA TTC GAT TTT AGC AAA TG
targeted_Scr_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTT CGA ATG CAA CTT GTT CTG C
targeted_Lab_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTC CCT GAT AAT GGC GAA CAG T
targeted_Antp_IS_fw	This study	AAG CAG TGG TAT CAA CGC AGA GTA GAG GAA CAG CAA AGC GAA A
targeted_Dfd_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTT CGG ATT GTT GCT GTT GAA G
targeted_Ubx_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTC AGA ATT TTG CTC GCA TTC A
targeted_AbdA_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTC ATG CGT TGC TCT ATC AAA
targeted_AbdB_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTA ATA TAA TGC TCG GGG CAA A
targeted_Scr_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTA TTG GGC GAT ACA AAC GAA G
targeted_Lab_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTC CCT TCA ACT TTG CTT GCT C
targeted_Antp_IS_rv	This study	AAG CAG TGG TAT CAA CGC AGA GTA ACC ATA CCC AGT CCA CCA A
qPCR_act5C_fw	This study	AGG AGG AGG AGG AGA AGT CG
qPCR_act5C_rv	This study	TGT GCT GCA CTC CAA ACT TC

Reagents and Tools table (continued)

Reagent/Resource	Reference or source	Identifier or Catalog Number
qPCR_Dfd_fw	Hu *et al* (2013)	GCG AAC GGA TCA TCT ACC CC
qPCR_Dfd_rv	Hu *et al* (2013)	ATC TGA TGG CGT GTG TAG GC
qPCR_DIP-kappa_fw	Hu *et al* (2013)	ATC CCC TCA AAG GGA AAA CAC A
qPCR_DIP-kappa_rv	Hu *et al* (2013)	GAA TCG CGG AAA GTC GGA ATC
qPCR_DIP-gamma_fw	Hu *et al* (2013)	AAC CAG CAT CAC GAG AGC AG
qPCR_DIP-gamma_rv	Hu *et al* (2013)	GCC GGA TAT GTG ACG TTG TTG
qPCR_exd_fw	Hu *et al* (2013)	CGG AGC AAT CAC TTG ACG AGG
qPCR_exd_rv	Hu *et al* (2013)	CGA GAG GAC GGT CTT CTC CTT
Chemicals, enzymes and other reagents		
5xSMART First Strand buffer	Clonetech	N/A
Ampure XP beads	Beckman	A63880
DNase I (RNase free)	Invitrogen	AM222
Formaldehyde	Sigma Aldrich	252549
KAPA HiFi HotStart ReadyMix	fisherscientific	NC0295239
oligo (dT)18	Thermo Scientific	S0132
RevertAid First Strand cDNA synthesis Kit	Thermo Scientific	K1622
RNAse Inhibitor	Thermo Scientific	EO0381
SMARTScribe™ Reverse Transcriptase	clonetech	639537
SSC (20X)	Sigma Aldrich	15557044
SYBR green for qPCR	Invitrogen	117733-038
Tn5	PepCore EMBL	N/A
TRIzol reagent	Ambion by Life Technologies	15596026
VectaShield + DAPI	Vector Laboratories	H-1200
HCR probe *Vglut*_B1	Molecular Instruments	NM_001273010.1
HCR probe *DIP-gamma*_B2	Molecular Instruments	NM_143392
HCR probe *DIP-kappa*_B3	Molecular Instruments	NM_165049, NM_001298975.1
HCR probe *otk*_B4	Molecular Instruments	NM_078981.3, NM_001299396.1
HCR probe *Lim1*_B4	Molecular Instruments	NM_132277.3, NM_001272429.2
HCR probe *unc4*_B5	Molecular Instruments	NM_001298460.1, NM_133014.3, NM_001298461.1
HCR probe *twit*_B2	Molecular Instruments	NM_136223
HCR probe *Vmat*_B3	Molecular Instruments	NM_001014524.2, NM_001014525.3, NM_001014526.2, NM_001274027.2
Software		
Fiji (ImageJ)	Schindelin *et al* (2012)	https://imagej.net/Fiji
IndeXplorer	Velten *et al* (2017)	https://git.embl.de/velten/indeXplorer
iRegulon	Janky *et al* (2014)	http://iregulon.aertslab.org/
Other		
Illumina NextSeq platform	Illumina	
qTOWER3	Analytik Jena	
Leica TCS SP8 confocal microscope	Leica	

Methods and Protocols

Drosophila strains and experimental crosses

We used the *OK371*-GAL4 driver (Mahr & Aberle, 2006) crossed to UAS-*mRFP* to perform single cell sorting of MNs and the *Mhc-TAU-GFP* line (Chen & Olson, 2001) for sorting of somatic muscle cells. Crosses were kept for 1 h at 25°C to oviposit on apple juice plates with yeast paste. Subsequently, eggs were incubated on apple juice plates for additional 19 h, after which embryos were dissociated for FACS sorting.

In order to validate protein expression of specific marker genes, we used the following endogenously GFP-tagged fly lines: *frq1::GFP* (BL:60284), *mirr::GFP* (BL:68183) and *lab::GFP* (BL:66551) lines.

For genetic experiments, we used the pan-neural *elav*-GAL4 driver (Luo *et al*, 1994) and the muscle-specific *Mef2*-GAL4 driver (Ranganayakulu *et al*, 1998) as main driver. In addition, we combined these two driver lines (*elav*-GAL4; *Mef2*-GAL4 driver) for this study. To confirm phenotypes, in MNs, the *OK6*-GAL4 driver (Aberle *et al*, 2002) was used in a few examples. In this study, we used second-generation *TRiP RNAi* lines from Bloomington (Ni *et al*, 2011), as we found that these *short hairpin RNAs* (*shRNAs*) are more effective in embryonic stages. Therefore, only one UAS-*RNAi* line fulfilling the required criteria has been available for most of the target genes. The UAS-*Dfd^{RNAi}* was designed for this study using the second generation of TRiP system. The UAS-*mirr^{RNAi}* (*BL:65877*), UAS-*hth^{RNAi}* (*BL:34637*), UAS-*DIPgamma^{RNAi}* (*BL:80461*), UAS-*DIPkappa^{RNAi}*(*BL:31740*) were obtained from the Bloomington Stock Center. The UAS overexpression lines UAS-*Ubx*-HA (Domsch *et al*, 2019) and AS-*Dfd*-HA were designed by Katrin Domsch as described in (Domsch *et al*, 2019), the UAS-lab and UAS-*dpr1* (BL:31740) lines were from the Bloomington Stock Center, and the UAS-*DIPgamma* (F003086) and UAS-*dpr11* (F003032) lines were from the FlyORF collection. For innervation rate assays, we crossed virgins of the corresponding driver line to males carrying UAS-*RNAi* or UAS overexpression constructs. As control, we performed the experiment under the same conditions with UAS-*mRFP* lines or UAS-*EGFP^{RNAi}* lines. The crosses were kept for at least 16 h at 29°C for knockdown experiments and 20 h at 25°C for overexpression experiments on apple juice plates. Early-stage 17 embryos were selected after embryo fixation.

For hatching rate assays, the *elav*-GAL4 driver and *Mef2*-GAL4 drivers were crossed to UAS-*RNAi* lines. UAS-*mRFP* lines or UAS-*EGFP^{RNAi}* lines crossed to the above-mentioned driver lines were used as controls. Crosses were set up in duplicates (equal amounts of males and females for every replicate, sample and control), and egg laying was performed at 29°C for 3 h on apple-juice plates with yeast paste. Subsequently, eggs were washed, counted and unfertilized eggs were removed or counted and transferred on fresh apple juice plates without yeast paste. Eggs were incubated for additional 24 h and then the hatching rate was quantified.

Plasmid construction, transgenesis and antibody production

The UAS-*Dfd^{RNAi}* line was generated following the conventional TRiP protocol (second generation).

UAS-HA-*Dfd* fly line: The full *Dfd* coding region was cloned into the pUAS-attB vector (Bischof *et al*, 2007) using a forward primer with an *Eco*RI restriction site and an HA tag sequence as well as a reverse primer with an *Xba*I restriction site. The UAS-*Dfd* construct was injected by BestGene into attP5 embryos (second chromosome). Primers and sequence maps are available upon request. Dfd antibody: The Dfd antibody was generated using the pGEX-purification system (GElifesciences). The open reading frame of *Dfd* was cloned in the pGEX-6P-2 vector using *Eco*RI and *Xho*I restriction site. The protein was purified according to the protocol (GElifesciences) and eluted by using the PreScission Protease site. The immunization and antibody handling were performed by the Charles Rivers Company.

Immunohistochemistry

The embryos were fixed by bleaching with 100% Clorox first for 2 min to remove the chorion. After washing in water, embryos were transferred to fixing solution (3.7% formaldehyde in PBS + 100% heptane) and incubated for 20 min at room temperature on a nutator. Fixing was stopped by removal of formaldehyde. Then, equal amounts of methanol were added to heptane and vortexed for about 40 s to remove the vitellin membrane. Subsequently, the heptane phase was removed, and embryos were washed in methanol.

For antibody stainings, embryos were hydrated and washed for three times in PBT (with 0.1% Tween 20). The primary antibodies were used at 4°C overnight from Abcam (rat anti-Myosin 1:1,000) DSHB (mouse anti-FasII, 1:50; mouse anti-Scr, 1:50; mouse anti-Abd A, 1:50; mouse anti-Antp; rat anti-Elav 1:50, mouse anti-AbdB 1:50), Invitrogen (rabbit anti-GFP, 1:500), provided by Katrin Domsch (guinea pig anti-Dfd 1:500), guinea pig anti-Ubx (1:500) (Domsch *et al*, 2019), rabbit anti-Lab was given by T. Kaufmann, rabbit anti-Hth provide by R. Mann and the rat anti-Vvl was given by the Johnson lab. After 3× washing in PBT (with 0.1% Tween 20), embryos were incubated for 2 h at room temperature with secondary antibodies from Jackson Immunoresearch. Vectashield with DAPI or TSO was used as mounting medium.

For double stainings with antibodies originating from the same animal (mouse anti-FasII + mouse anti-Antp; mouse anti-FasII + mouse anti-Abd A; mouse anti-FasII + mouse anti-Scr), we performed a modified protocol for sequential antibody staining by TSA (according to the manufacture protocol). Primary and secondary antibodies were used with similar concentrations than for antibody staining's of embryos.

In situ *hybridization chain reaction (HCR)*

Multiplex *in situ* hybridization experiments were performed on fixed late-stage *Drosophila* embryos according to the HCR version 3.0 protocol (Choi *et al*, 2018) and personal communication with Christoph Schaub. Customized probes were generated for this study by Molecular Instruments targeting all isoforms of each target gene (*VGlut* B1, *DIP-gamma* B2, *DIP-kappa* B3, *otk* B3, *Lim1* B4, *unc4* B5, *twit* B2, *Vmat* B3). A maximum of three targets were multiplexed in the same *in situ* hybridization experiment or the *in situ* hybridization protocol was combined with the immunohistochemistry protocol.

In brief, late-stage embryos were collected, bleached and fixed in 3.7% formaldehyde in PBS + 100% heptane. Then, the vitelline membrane of embryos was removed by vortexing and the embryos were dehydrated in methanol. For *in situ* hybridization experiments, embryos were first prehybridized and then hybridized at 37°C with different combinations of customized probes with corresponding adaptors in 30% probe hybridization buffer (30% formamide, 5xSSC, 9 mM citric acid pH6, 0.1% Tween20, 50 µg/ml heparin, 1× Denhardt's solution, 5% dextran sulphate). After washing in 5× SSCT (5× SSC, 0.1% Tween20, 5% dextran sulphate), fluorescently labelled hairpins were pre-amplified for each linker (B1, B2, B3, B4, B5) in amplification buffer (5xSSC, 0.1% Tween20, 5% dextran sulphate). Prepared linkers were incubated with the embryos at room temperature. After washing, samples were either mounted for microscopy or an additional immunohistochemistry protocol followed as described in the section above.

Quantitative PCR

Early-stage 17 embryos were subjected to RNA extraction using TRIzol. RNA was then isolated by chloroform/isopropanol extraction. After DNA digestion (DNase I, Invitrogen), 1 μg of total RNA was converted to cDNA using the RevertAid First Strand cDNA synthesis Kit (Invitrogen) and oligo(dT)$_{18}$ primers (Fisher Scientific). Quantitative PCR was performed using the SYBR Green Master Mix (Invitrogen), 150 nM primer, in 96-well plates, conducted on the qTOWER3 machine (Analytik Jena).

Microscopy and image analysis

Fixed embryos were imaged on a Leica TCS SP8 confocal microscope. The 20× objective was used for imaging of embryos and the 63× objective to visualize the neuromuscular junction of the MHE muscle. For homeodomain TF cluster analysis, we used as reference the highly stereotypic neuronal projection from the MN2a to MHE and the MN3 to MHD labelled by the motoneuronal marker FasII. Going from MN2a, we identify the upper (MN2c) and lower (MN4) MNs based on the location. Images were processed by Fiji and the integrated ROI manager was used to label and save locations of cells and projection based on FasII, these labels were transferred across pictures. To enable precise localization across different confocal images, we used the Fijiyama plugin for 3D alignment. Embryos used for quantification of innervation rate assays were processed with a standardized imaging pipeline to facilitate counting (programed by Patrick van Nierop Y Sanchez, see https://github.com/patrickvannieropysanchez/Velten_et_al._2021). Confocal pictures shown in figures were processed manually to optimize the result.

For AP axis measurements along the ventral nerve cord, we used FasII staining as reference and measured protein intensities of Lab, Dfd, Scr, Antp, Ubx, Abd A and Abd B by Fiji. We normalized the length of the ventral nerve cord between different embryos (0 = anterior to 2000 = posterior), set thresholds for the protein intensities to remove the overall unspecific background caused by antibody stainings and normalized the protein intensities of different the antibodies (0=min intensity to 100=max intensity). For counting motoneuron numbers, 20x resolution was used to orientate the embryos with help from the Fiji plugin "clear volume", cells were counted at 63× resolution using the Fiji multiple point tool. Vnd antibody staining was used to identify segment boundaries in the embryonic ventral nerve cord (McDonald *et al*, 1998).

Dissociation of embryonic cells and flow cytometry

Collections of 19- to 20-h-old *Drosophila* embryos were dechorionated with bleach (5% Chloride) for 2 min. For dissociation, embryos were transferred into syringes filled with 1ml of 10x Trypsin-EDTA (0.5% Trypsin). For additional mechanical dissociation, embryos were transferred 20x between two syringes (25G needle) and 10× between two syringes with 27G needles. Debris was reduced by filtering through a 35-μm cell strainer. The cells were stained for 5 min with DRAQ5 (1 mg/ml) and DAPI (1 mg/ml). Subsequently, cells were sorted into 96-well plates containing 5 μl of Smart-Seq2 lysis buffer at 4°C by a BD FACS Aria III flow cytometer equipped with 405nm, 488nm, 561nm and 633nm laser. Directly after cell sorting, plates were shock-frozen in liquid nitrogen.

Single-cell transcriptome sequencing

A pooled cell population of *OK371 > RFP*-positive motoneuronal cells and a pooled population of *Mhc-TAU-GFP*-positive somatic muscle cells derived from FACS-sorted 19- to 20-h-old *Drosophila* embryos were used for scRNA sequencing. The standard Smart-Seq2 protocol (Picelli *et al*, 2014) was modified by targeting low-expressed *Hox* genes (*lab, Dfd, Scr, Antp, Ubx, abdA, AbdB*) during the reverse transcription and PCR stages. Therefore, 1μM targeted primer mix of each *Hox gene* (see key resource table) was added to the PCR and RT buffers, respectively. The above-described approach targets all expressed isoforms of each *Hox* gene. Method-specific biases were ruled out by quality checks via Bioanalyzer of the whole transcriptome after Smart-Seq2 preparation. Libraries of scRNA transcriptomes were created using a home-made Tn5 transposase (Hennig *et al*, 2018) and seventeen 96-well plates were sequenced with 75bp single-end on an Illumina NextSeq platform.

Raw data processing, quality control and normalization

Sequencing reads were demultiplexed and the poly-A tail trimmed. Read count tables were generated by pseudo-alignment to the cDNA of the *Drosophila* transcriptome (BDGP6 ensemble) using Kallisto (Bray *et al*, 2016). For quality control, we used standard settings (Velten *et al*, 2017): cells were removed unless they contained at least 10 reads for each of at least 500 genes, and genes were removed unless they were expressed in at least five cells with 10 reads each. For basic cell type discovery and PCA, data were normalized and scaled to an estimate of biological variance using the indeXplorer pipeline (Velten *et al*, 2017).

Such scaled and mean-centred expression values take very high values if the gene is expressed in few cells and low values if the gene is expressed in many cells (see Fig EV4G). Library-size normalized data emphasize the absolute expression level of the genes. Neither of the two strategies is ideal when aiming to identify a "code" using clustering strategies. Binarization of the gene expression levels would constitute a viable alternative (see for example (Li *et al*, 2017)) but cannot capture biologically meaningful quantitative differences between cells expressing the gene "highly" or "lowly". For the heatmap depicting the homeodomain TF expression, we therefore normalized the data gene-wise by

$$N_i = \log\left(k\frac{X_i}{\sum X} + 1\right)$$

where X is the vector of size factor normalized counts for a given gene. N approaches binarization for large k. Here, a value of $k = 10^6$ was chosen. This leads to a behaviour where normalized values are essentially binarized for lowly expressed genes, where quantitative differences cannot be captured in a meaningful manner; by contrast, for highly expressed genes, quantitative expression levels play a role (see Fig EV4G). We have previously used similarly motivated strategy for data normalization in the context of predicting cell–cell interactions using scRNA-Seq data (Baccin *et al*, 2020).

Clustering and dimensionality reduction

Following the indeXplorer pipeline, PCA was computed using scaled, normalized expression values of all genes. t-SNE was then computed on the first 10 principal components based on a visual

inspection of the PCA Elbow plot. Alternatively, and following (Li *et al*, 2017), cells were clustered by hierarchical clustering using only the 20 most variably expressed genes of our scRNA dataset; Ward linkage was used on an Euclidean distance metric. The *twit*^low cluster was selected for further analysis on variability in gene expression. To this end, we used the statistical method developed by (Brennecke *et al*, 2013) to identify variable genes. The variable homeodomain TFs thereby identified were normalized as described above and used as input for unsupervised hierarchical clustering using ward linkage and Euclidean distance matrix to avoid that clustering is affected by technical noise. To validate this point, we showed that homeo-clusters are not significantly associated with technical covariability (sequencing depth) by Wilcoxon test (Fig EV4H). Then, we split the clusters in a supervised manner according to the structure of the dendrogram and compared the approximate number of cells labelled by the *OK371*-GAL4 driver during this stage with the estimated cluster number.

Inference of spatial position from single-cell gene expression data

To infer spatial position from single-cell gene expression data, we first created a reference map of protein expression for seven *Hox* genes (Fig EV4A and B). To that end, immunofluorescence experiments were performed as described above.

Immunofluorescence data for gene g were thereby represented as a function of position $Y_g(x) \in (0, 1)$, whereas single-cell gene expression was represented as a matrix of read count values across genes and cells, $D_{g,c} \in \mathbb{N}_0$. Cells not expressing any Hox gene were excluded. We then assumed that the probability of observing $D_{g,c}$ given that the position of cell c is really x_c depends on the expression of g at that position:

$$p(D_{g,c}|x_c) \sim \begin{cases} Pois(D_{g,c}|r_g * Y_g(x)) & \text{if } Y_g(x_c) > 0 \\ Pois(D_{g,c}|\lambda) & \text{if } Y_g(x_c) = 0 \end{cases}$$

here λ is a constant corresponding to a background number of reads observed in non-expressing cells, and r_g is a gene-wise scaling factor that estimates the average number of RNA-seq reads in a cell maximally expressing the protein. A maximum likelihood estimate of x_c is then obtained by computing

$$\hat{x}_c = \arg\max_x \prod_g p(D_{c,g}|x)$$

For the analysis shown in the manuscript, λ was set to 0.1 (Kharchenko *et al*, 2014) and r_g was set to the mean gene expression in expressing cells:

$$r_g = \frac{\sum_c D_{c,g}}{\sum_c D_{c,g} > 0}$$

We used Latin Hypercube sampling across a wide range of values for λ and r to demonstrate that the position estimate was independent of parameter choice.

To identify genes with spatially variable expression, B-Spline models with 3 degrees of freedom were fitted through scaled and normalized gene expression values for each gene individually, using inferred position as the only covariate. These models were then compared to null models not containing the position term using the

Bayesian information criterion, similar to the workflow for selecting genes with variable expression over pseudotime described in (Velten *et al*, 2017).

For visualization and clustering, expression values for all variably expressed genes were arranged by inferred position and a floating mean was computed by 1D-convolution with an absolute exponential kernel with decay rate 10. Smoothened gene expression values obtained thereby were then compared to immunofluorescence images for model validation (Fig EV4D) or used for clustering of gene into modules with coherent expression patterns over space (Fig 1D and E).

ZINB-WaVE analysis

To identify gene whose expression was variable but statistically independent from the AP axis, we made use of the ZINB-WaVE model (Risso *et al*, 2018, 2019). Unlike a PCA, ZINB-WaVE separately estimates a matrix of known-covariate factors as well as a matrix of unknown-covariate factors; furthermore, ZINB-WaVE uses a zero-inflated negative binomial distribution to account for the sparse nature of single-cell RNA-seq data. We ran ZINB-WaVE using default settings and the number of genes observed as well as the inferred AP position as known sample-level covariates. While the PCA of the dataset was dominated by effects related to the number of genes observed, viability dye incorporation, and AP position (Fig EV4E), the unknown-covariate factors from ZINB-WaVE arranged genes according to their dorsal–ventral expression (Fig 1F). Two of the predicted DV markers were validated by Immunofluorescence (Fig EV4F).

ChIP-Seq reanalysis and iRegulon analysis

To identify the relationship between motoneuronal genes expressed in our scRNA-Seq dataset and genes bound by the homeodomain TF Dfd, we used a whole embryo Dfd ChIP (Sorge *et al*, 2012) to investigate if Ig domain proteins are overrepresented among putative Dfd-regulated genes compared to other genes expressed in MNs. Genes were classified as regulated by Dfd when Dfd binding to the target gene was detected max. 1 kb upstream of the promotor region (Sorge *et al*, 2012). To calculate a *P*-value for enrichment, we performed a hypergeometric test. The Ubx ChIP-Seq data were analysed as described in (Domsch *et al*, 2019). To illustrate the overlap of neuronal Ubx-bound genes with the genes from the Ig expression maps, mapped neuronal Ubx peaks were assigned to genes in close proximity and these genes were used in the Venn diagram to illustrate the common and unique populations.

The iRegulon (Janky *et al*, 2014) analysis was used to identify TF motifs enriched in the vicinity of Ig-encoding genes expressed in our motoneuronal scRNA dataset. Therefore, we performed an iRegulon analysis on all Ig domain encoding proteins expressed in MNs under standard settings (9,713PWM; 5 kb upstream, 50 UTR and first intron with standard cut-off) to identify the 15 highest ranked TF motifs. We next used iRegulon to investigate which TF families are predicted to bind to these motifs and chose three of the five motifs predicted to be regulated by homeodomain TFs.

Data visualization

All plots were generated using the ggplot2 (v. 3.2.1) and pheatmap (v. 1.0.12) packages in R 3.6.2. Boxplots are defined as follows: The

middle line corresponds to the median; lower and upper hinges correspond to first and third quartiles. The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.

Data and software availability

Single-cell transcriptomics raw data and count tables are available from gene expression omnibus, ID: GSE155578 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155578for the motoneuronal dataset) and ID: GSE155586 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155586; for the muscle dataset).

Most analyses were done using the indeXexplorer software for interactive exploration of single-cell RNA-seq datasets (Velten *et al*, 2017), which is available from https://git.embl.de/velten/indeXplorer. Custom scripts for inference of spatial position are available from the corresponding authors upon request.

Expanded View for this article is available online.

Acknowledgements

Author contributions

Disclosure and competing interests statement

References

Aberle H, Haghighi AP, Fetter RD, McCabe BD, Magalhães TR, Goodman CS (2002) Wishful thinking encodes a BMP type II receptor that regulates synaptic growth in Drosophila. *Neuron* 33: 545–558

Allan DW, Thor S (2015) Transcriptional selectors, masters, and combinatorial codes: regulatory principles of neural subtype specification. *Wiley Interdiscip Rev Dev Biol* 4: 505–528

Allen AM, Neville MC, Birtles S, Croset V, Treiber CD, Waddell S, Goodwin SF (2020) A single-cell transcriptomic atlas of the adult *Drosophila* ventral nerve cord. *Elife* 9: e54074

Anderson MG, Perkins GL, Chittick P, Shrigley RJ, Johnson WA (1995) Drifter, a Drosophila POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia. *Genes Dev* 9: 123–137

Angelini DR, Kaufman TC (2005) Comparative developmental genetics and the evolution of arthropod body plans. *Annu Rev Genet* 39: 95–119

Arber S (2012) Motor circuits in action: specification, connectivity, and function. *Neuron* 74: 975–989

Ashley J, Sorrentino V, Lobb-Rabe M, Nagarkar-Jaiswal S, Tan L, Xu S, Xiao Q, Zinn K, Carrillo RA (2019) Transsynaptic interactions between IgSF proteins DIP-α and Dpr10 are required for motor neuron targeting specificity. *Elife* 8: e42690

Baccin C, Al-Sabah J, Velten L, Helbling PM, Grünschläger F, Hernández-Malmierca P, Nombela-Arrieta C, Steinmetz LM, Trumpp A, Haas S (2020) Combined single-cell and spatial transcriptomics reveals the molecular, cellular and spatial bone marrow niche organization. *Nat Cell Biol* 22: 38–48

Bageritz J, Willnow P, Valentini E, Leible S, Boutros M, Teleman AA (2019) Gene expression atlas of a developing tissue by single cell expression correlation analysis. *Nat Methods* 16: 750–756

Bhat KM (1999) Segment polarity genes in neuroblast formation and identity specification during *Drosophila* neurogenesis. *BioEssays* 21: 472–485

Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific φC31 integrases. *Proc Natl Acad Sci USA* 104: 3312–3317

Bossing T, Udolph G, Doe CQ, Technau GM (1996) The embryonic central nervous system lineages of *Drosophila melanogaster* I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Dev Biol* 179: 41–64

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34: 525–527

Brennecke P, Anders S, Kim JK, Kołodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC *et al* (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* 10: 1093–1095

Broihier HT, Skeath JB (2002) *Drosophila* homeodomain protein dHb9 directs neuronal fate via crossrepressive and cell-nonautonomous mechanisms. *Neuron* 35: 39–50

Broihier HT, Kuzin A, Zhu Y, Odenwald W, Skeath JB (2004) *Drosophila* homeodomain protein Nkx6 coordinates motoneuron subtype identity and axonogenesis. *Development* 131: 5233–5242

Carrillo RA, Özkan E, Menon KP, Nagarkar-Jaiswal S, Lee P-T, Jeon M, Birnbaum ME, Bellen HJ, Garcia KC, Zinn K (2015) Control of synaptic connectivity by a network of *Drosophila* IgSF cell surface proteins. *Cell* 163: 1770–1782

Certel SJ, Thor S (2004) Specification of *Drosophila* motoneuron identity by the combinatorial action of POU and LIM-HD factors. *Development* 131: 5429–5439

Chen EH, Olson EN (2001) Antisocial, an intracellular adaptor protein, is required for myoblast fusion in *Drosophila*. *Dev Cell* 1: 705–715

Cheng S, Ashley J, Kurleto JD, Lobb-Rabe M, Park YJ, Carrillo RA, Özkan E (2019) Molecular basis of synaptic specificity by immunoglobulin superfamily receptors in *Drosophila*. *Elife* 8: e41028

Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, Pierce NA (2018) Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145: dev165753

Couto A, Alenius M, Dickson BJ (2005) Molecular, anatomical, and functional organization of the *drosophila* olfactory system. *Curr Biol* 15: 1535–1547

Couton L, Mauss AS, Yunusov T, Diegelmann S, Evers JF, Landgraf M (2015) Development of connectivity in a motoneuronal network in *Drosophila* Larvae. *Curr Biol* 25: 568–576

Dasen JS, Jessell TM (2009) Chapter Six Hox Networks and the origins of motor neuron diversity. *Curr Top Dev Biol* 88: 169–200

Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, Aibar S, Makhzami S, Christiaens V, Bravo González-Blas C *et al* (2018) A single-cell transcriptome atlas of the aging *Drosophila* Brain. *Cell* 174: 982–998.e20

Demireva EY, Shapiro LS, Jessell TM, Zampieri N (2011) Motor neuron position and topographic order imposed by β- and γ-catenin activities. *Cell* 147: 641–652

Diederich RJ, Merrill VK, Pultz MA, Kaufman TC (1989) Isolation, structure, and expression of labial, a homeotic gene of the Antennapedia Complex involved in Drosophila head development. *Genes Dev* 3: 399–414

Domsch K, Carnesecchi J, Disela V, Friedrich J, Trost N, Ermakova O, Polychronidou M, Lohmann I (2019) The Hox transcription factor Ubx stabilizes lineage commitment by suppressing cellular plasticity in *Drosophila*. *Elife* 8: e42675

Ekker SC, Jackson DG, von Kessler DP, Sun BI, Young KE, Beachy PA (1994) The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. *Embo J* 13: 3551–3560

Friedrich J, Sorge S, Bujupi F, Eichenlaub MP, Schulz NG, Wittbrodt J, Lohmann I (2016) Hox function is required for the development and maintenance of the *Drosophila* feeding motor unit. *Cell Rep* 14: 850–860

Garces A, Thor S (2006) Specification of *Drosophila* aCC motoneuron identity by a genetic cascade involving even-skipped, grain and zfh1. *Development* 133: 1445–1455

Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, Sopp P, Norfo R, Rodriguez-Meira A, Ashley N *et al* (2017) Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat Med* 23: 692–702

Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, Besir H, Remans K, Steinmetz LM (2018) Large-Scale Low-Cost NGS library preparation using a robust Tn5 purification and tagmentation protocol. *G3* 8: 79–89

Hessinger C, Technau GM, Rogulja-Ortmann A (2016) The *Drosophila* Hox gene ultrabithorax acts in both muscles and motoneurons to orchestrate formation of specific neuromuscular connections. *Development* 144: 139–150

Hoang B, Chiba A (2001) Single-cell analysis of *Drosophila* larval neuromuscular synapses. *Dev Biol* 229: 55–70

Hobert O (2021) Homeobox genes and the specification of neuronal identity. *Nat Rev Neurosci* 22: 627–636

Hu Y, Sopko R, Foos M, Kelley C, Flockhart I, Ammeux N, Wang X, Perkins L, Perrimon N, Mohr SE (2013) FlyPrimerBank: an online database for Drosophila melanogaster gene expression analysis and knockdown evaluation of RNAi reagents. *G3* 3: 1607–1616

Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D *et al* (2014) iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *Plos Comput Biol* 10: e1003731

Jurata LW, Thomas JB, Pfaff SL (2000) Transcriptional mechanisms in the development of motor control. *Curr Opin Neurobiol* 10: 72–79

Keleman K, Dickson BJ (2001) Short- and long-range repulsion by the *Drosophila* Unc5 netrin receptor. *Neuron* 32: 605–617

Kharchenko PV, Silberstein L, Scadden DT (2014) Bayesian approach to single-cell differential expression analysis. *Nat Methods* 11: 740–742

Kim MD, Wen Y, Jan Y-N (2009) Patterning and organization of motor neuron dendrites in the *Drosophila* larva. *Dev Biol* 336: 213–221

Kim NC, Marqués G (2012) The Ly6 neurotoxin-like molecule target of wit regulates spontaneous neurotransmitter release at the developing neuromuscular junction in *Drosophila*. *Dev Neurobiol* 72: 1541–1558

Labrador JP, O'Keefe D, Yoshikawa S, McKinnon RD, Thomas JB, Bashaw GJ (2005) The homeobox transcription factor even-skipped regulates netrin-receptor expression to control dorsal motor-axon projections in *Drosophila*. *Curr Biol* 15: 1413–1419

Landgraf M, Bossing T, Technau GM, Bate M (1997) The origin, location, and projections of the embryonic abdominal motoneurons of *Drosophila*. *J Neurosci* 17: 9642–9655

Landgraf M, Roy S, Prokop A, VijayRaghavan K, Bate M (1999) Even-skipped determines the dorsal growth of motor axons in *Drosophila*. *Neuron* 22: 43–52

Landgraf M, Jeffrey V, Fujioka M, Jaynes JB, Bate M (2003) Embryonic origins of a motor system: motor dendrites form a myotopic map in *Drosophila*. *Plos Biol* 1: e41

Landgraf M, Thor S (2006) Development of *Drosophila* motoneurons: specification and morphology. *Semin Cell Dev Biol* 17: 3–11

Li H, Horns F, Wu B, Xie Q, Li J, Li T, Luginbuhl DJ, Quake SR, Luo L (2017) Classifying *Drosophila* olfactory projection neuron subtypes by single-cell RNA sequencing. *Cell* 171: 1206–1220.e22

Li H, Li T, Horns F, Li J, Xie Q, Xu C, Wu B, Kebschull JM, McLaughlin CN, Kolluru SS *et al* (2020) Single-cell transcriptomes reveal diverse regulatory strategies for olfactory receptor expression and axon targeting. *Curr Biol* 30: 1189–1198.e5

Luecken MD, Theis FJ (2019) Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 15: e8746

Luo L, Liao YJ, Jan LY, Jan YN (1994) Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Gene Dev* 8: 1787–1802

Mahr A, Aberle H (2006) The expression pattern of the *Drosophila* vesicular glutamate transporter: a marker protein for motoneurons and glutamatergic centers in the brain. *Gene Expr Patterns* 6: 299–309

McDonald JA, Holbrook S, Isshiki T, Weiss J, Doe CQ, Mellerick DM (1998) Dorsoventral patterning in the *Drosophila* central nervous system: the

vnd homeobox gene specifies ventral column identity. *Gene Dev* 12: 3603–3612

McKay DJ, Estella C, Mann RS (2009) The origins of the Drosophila leg revealed by the cis-regulatory architecture of the Distalless gene. *Development* 136: 61–71

McLaughlin CN, Brbić M, Xie Q, Li T, Horns F, Kolluru SS, Kebschull JM, Vacek D, Xie A, Li J *et al* (2021) Single-cell transcriptomes of developing and adult olfactory receptor neurons in *Drosophila. Elife* 10: e63856

Meng JL, Heckscher ES (2020) Development of motor circuits: from neuronal stem cells and neuronal diversity to motor circuit assembly. *Curr Top Dev Biol* 142: 409–442

Nakamura M, Baldwin D, Hannaford S, Palka J, Montell C (2002) Defective proboscis extension response (DPR), a member of the ig superfamily required for the gustatory response to salt. *J Neurosci* 22: 3463–3472

Nassif C, Noveen A, Hartenstein V (1998) Embryonic development of the *Drosophila* brain. I. Pattern of pioneer tracts. *J Comp Neurol* 402: 10–31

Ni J-Q, Zhou R, Czech B, Liu L-P, Holderbaum L, Yang-Zhou D, Shim H-S, Tao R, Handler D, Karpowicz P *et al* (2011) A genome-scale shRNA resource for transgenic RNAi in *Drosophila. Nat Methods* 8: 405–407

Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* 133: 1277–1289

Özel MN, Simon F, Jafari S, Holguera I, Chen Y-C, Benhra N, El-Danaf RN, Kapuralin K, Malin JA, Konstantinides N *et al* (2021) Neuronal diversity and convergence in a visual system developmental atlas. *Nature* 589: 88–95

Özkan E, Carrillo RA, Eastman CL, Weiszmann R, Waghray D, Johnson KG, Zinn K, Celniker SE, Garcia KC (2013) An Extracellular interactome of immunoglobulin and LRR proteins reveals receptor-ligand networks. *Cell* 154: 228–239

Philippidou P, Walsh C, Aubin J, Jeannotte L, Dasen JS (2012) Sustained Hox5 gene activity is required for respiratory motor neuron development. *Nat Neurosci* 15: 1636–1644

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9: 171–181

Ranganayakulu G, Elliott DA, Harvey RP, Olson EN (1998) Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. *Dev Camb Engl* 125: 3037–3048

Reilly MB, Cros C, Varol E, Yemini E, Hobert O (2020) Unique homeobox codes delineate all the neuron classes of *C. elegans. Nature* 584: 595–601

Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, Meer EJ, Terry JM, Riordan DP, Srinivas N *et al* (2020) Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol* 38: 954–961

Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert J-P (2018) A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun* 9: 284

Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert J-P (2019) Publisher Correction: a general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun* 10: 646

Sanes JR, Zipursky SL (2020) Synaptic specificity, recognition molecules, and assembly of neural circuits. *Cell* 181: 1434–1435

di Sanguinetto SADT, Dasen JS, Arber S (2008) Transcriptional mechanisms controlling motor neuron diversity and connectivity. *Curr Opin Neurobiol* 18: 36–43

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9: 676–682

Schmidt H, Rickert C, Bossing T, Vef O, Urban J, Technau GM (1997) The embryonic central nervous system lineages of *Drosophila melanogaster. Dev Biol* 189: 186–204

Schoofs A, Hanslik U, Niederegger S, Heinzel H, Spieß R (2010a) The thoracic muscular system and its innervation in third instar *Calliphora vicina* Larvae. II. Projection patterns of the nerves associated with the pro- and mesothorax and the pharyngeal complex. *J Morphol* 271: 969–979

Schoofs A, Niederegger S, van Ooyen A, Heinzel H-G, Spieß R (2010b) The brain can eat: establishing the existence of a central pattern generator for feeding in third instar larvae of *Drosophila* virilis and *Drosophila melanogaster. J Insect Physiol* 56: 695–705

Siegler MVS, Jia XX (1999) Engrailed negatively regulates the expression of cell adhesion molecules connectin and neuroglian in embryonic *Drosophila* nervous system. *Neuron* 22: 265–276

Sink H, Whitington PM (1991) Location and connectivity of abdominal motoneurons in the embryo and larva of *Drosophila melanogaster. J Neurobiol* 22: 298–311

Skeath JB (1999) At the nexus between pattern formation and cell-type specification: the generation of individual neuroblast fates in the *Drosophila* embryonic central nervous system. *BioEssays* 21: 922–931

Sorge S, Ha N, Polychronidou M, Friedrich J, Bezdan D, Kaspar P, Schaefer MH, Ossowski S, Henz SR, Mundorf J *et al* (2012) The cis-regulatory code of Hox function in *Drosophila. Embo J* 31: 3323–3333

Sperry RW (1963) Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc Natl Acad Sci USA* 50: 703–710

Stagg SB, Guardiola AR, Crews ST (2011) Dual role for *Drosophila* lethal of scute in CNS midline precursor formation and dopaminergic neuron and motoneuron cell fate. *Development* 138: 2171–2183

Sugino K, Clark E, Schulmann A, Shima Y, Wang L, Hunt DL, Hooks BM, Tränkner D, Chandrashekar J, Picard S *et al* (2019) Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. *eLife* 8: e38619

Sweeney LB, Chou Y-H, Wu Z, Joo W, Komiyama T, Potter CJ, Kolodkin AL, Garcia KC, Luo L (2011) Secreted semaphorins from degenerating larval ORN axons direct adult projection neuron dendrite targeting. *Neuron* 72: 734–747

Tan L, Zhang K, Pecot M, Nagarkar-Jaiswal S, Lee P-T, Takemura S-Y, McEwen J, Nern A, Xu S, Tadros W *et al* (2015) Ig superfamily ligand and receptor pairs expressed in synaptic partners in *Drosophila. Cell* 163: 1756–1769

Technau GM, Rogulja-Ortmann A, Berger C, Birkholz O, Rickert C (2014) Composition of a neuromere and its segmental diversification under the control of Hox genes in the embryonic CNS of *Drosophila. J Neurogenet* 28: 171–180

Thor S, Andersson SGE, Tomlinson A, Thomas JB (1999) A LIM-homeodomain combinatorial code for motor-neuron pathway selection. *Nature* 397: 76–80

Thor S, Thomas JB (2002) Motor neuron specification in worms, flies and mice: conserved and 'lost' mechanisms. *Curr Opin Genet Dev* 12: 558–564

Ulian-Benitez S, Bishop S, Foldi I, Wentzell J, Okenwa C, Forero MG, Zhu B, Moreira M, Phizacklea M, McIlroy G *et al* (2017) Kek-6: a truncated-Trk-like receptor for *Drosophila* neurotrophin 2 regulates structural synaptic plasticity. *Plos Genet* 13: e1006968

Urbach R, Volland D, Seibert J, Technau GM (2006) Segment-specific requirements for dorsoventral patterning genes during early brain development in *Drosophila*. *Development* 133: 4315–4330

Urbach R, Jussen D, Technau GM (2016) Gene expression profiles uncover individual identities of gnathal neuroblasts and serial homologies in the embryonic CNS of *Drosophila*. *Dev Camb Engl* 143: 1290–1301

van Oostrum M, Campbell B, Seng C, Müller M, tom Dieck S, Hammer J, Pedrioli PGA, Földy C, Tyagarajan SK, Wollscheid B (2020) Surfaceome dynamics reveal proteostasis-independent reorganization of neuronal surface proteins during development and synaptic plasticity. *Nat Commun* 11: 4990

Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, Hirche C, Lutz C, Buss EC, Nowak D *et al* (2017) Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol* 19: 271–281

Velten L, Story BA, Hernández-Malmierca P, Raffel S, Leonce DR, Milbank J, Paulsen M, Demir A, Szu-Tu C, Frömel R *et al* (2021) Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. *Nat Commun* 12: 1366

de Wit J, Ghosh A (2016) Specification of synaptic connectivity by cell surface interactions. *Nat Rev Neurosci* 17: 4

Xu S, Xiao QI, Cosmanescu F, Sergeeva AP, Yoo J, Lin Y, Katsamba PS, Ahlsen G, Kaufman J, Linaval NT *et al* (2018) Interactions between the Ig-superfamily proteins DIP-α and Dpr6/10 regulate assembly of neural circuits. *Neuron* 100: 1369–1384.e6

Xu C, Theisen E, Maloney R, Peng J, Santiago I, Yapp C, Werkhoven Z, Rumbaut E, Shum B, Tarnogorska D *et al* (2019) Control of synaptic specificity by establishing a relative preference for synaptic partners. *Neuron* 103: 865–877.e7

Yogev S, Shen K (2014) Cellular and molecular mechanisms of synaptic specificity. *Annu Rev Cell Dev Bi* 30: 1–21

Zarin AA, Asadzadeh J, Labrador J-P (2014) Transcriptional regulation of guidance at the midline and in motor circuits. *Cell Mol Life Sci* 71: 419–432

Zarin AA, Labrador J-P (2019) Motor axon guidance in *Drosophila*. *Semin Cell Dev Biol* 85: 36–47

Zarin AA, Mark B, Cardona A, Litwin-Kumar A, Doe CQ (2019) A multilayer circuit architecture for the generation of distinct locomotor behaviors in *Drosophila*. *Elife* 8: e51781

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G *et al* (2018) Molecular architecture of the mouse nervous system. *Cell* 174: 999–1014.e22

Zhu L, Lei J, Klei L, Devlin B, Roeder K (2018) Semisoft clustering of single-cell data. *Proc Natl Acad Sci USA* 116: 201817715