bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Molecular analysis of the midbrain dopaminergic niche during neurogenesis

Enrique M. Toledo[1], Gioele La Manno[1], Pia Rivetti di Val Cervo[1], Daniel Gyllborg[1], Saiful Islam[1,2], Carlos Villaescusa[1,3], Sten Linnarsson[1], Ernest Arenas[1].

[1]Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Scheeles väg 1, 17177, Stockholm, Sweden

[2]Present address: Department of Genetics, Stanford University, Stanford, USA

[3]Present address: Department of Molecular Medicine and Surgery. Center for Molecular Medicine. Karolinska Institutet

* Corresponding author: ernest.arenas@ki.se

Running title: The midbrain dopaminergic niche

Keywords: Weighted gene co-expression network analysis, RNA-sequencing, Fast Westfall-Young, Single-cell network deconvolution, Extracellular matrix

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Midbrain dopaminergic (mDA) neurons degenerate in Parkinson's disease and are one of the main targets for cell replacement therapies. However, a comprehensive view of the signals and cell types contributing to mDA neurogenesis is not yet available. By analyzing the transcriptome of the mouse ventral midbrain at a tissue and single-cell level during mDA neurogenesis we found that three recently identified radial glia types 1-3 (Rgl1-3) contribute to different key aspects of mDA neurogenesis. While Rgl3 expressed most extracellular matrix components and multiple ligands for various pathways controlling mDA neuron development, such as Wnt and Shh, Rgl1-2 expressed most receptors. Moreover, we found that specific transcription factor networks explain the transcriptome and suggest a function for each individual radial glia. A network controlling neurogenesis was found in Rgl1, progenitor maintenance in Rgl2 and the secretion of factors forming the mDA niche by Rgl3. Our results thus uncover a broad repertoire of developmental signals expressed by each midbrain cell type during mDA neurogenesis. Cells identified for their emerging importance are Rgl3, a niche cell type, and Rgl1, a neurogenic progenitor that expresses *ARNTL,* a transcription factor that we find is required for mDA neurogenesis.

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Our knowledge of the contribution of individual signals to midbrain dopaminergic (mDA) neuron development has grown in a considerable manner in recent years. It is currently thought that the development of mDA neurons is controlled by the interaction of multiple transcription factors and signaling pathways (Arenas *et al*, 2015; Smidt & Burbach, 2007). However, we still lack the systematic knowledge required to understand how all these factors are coordinated over time and space, in the complex signaling microenvironments in which the cells reside. These microenvironments, referred to as niches, consist of an extracellular matrix (ECM) and local paracrine/autocrine signals contributed by neighboring cells (Jones & Wagers, 2008). Far from static, such microenvironments are constantly being remodeled (Scadden, 2006; Jones & Wagers, 2008), leading to progressive changes in composition over time, which control fundamental developmental processes such as cell fate or the balance between proliferation and neurogenesis (Lu *et al*, 2011). In the nervous system, some of the critical cell types controlling neural development are radial glia cells. Typically, radial glia (Rgl) cells are bipolar cells with their somas located in the ventricular zone (VZ) and their processes extending to the ventricular cavity and radially to the pial surface. These cells, once thought to be a scaffold for neurogenesis, are currently thought to be transient stem cells, capable of limited self-renewal, of giving rise to different cell lineages and of undergoing neurogenesis by asymmetrical cell division (Götz & Huttner, 2005; Taverna *et al*, 2014). In the developing VM, it is unknown how many cells contribute to the DA neurogenic niche, how they contribute to neurogenesis, what factors do they express and what are the transcriptional networks operating in each of these cells. We previously reported that a proliferative floor plate Rgl cell gives rise to mDA neurons through the generation of a lineage with subsequent cellular intermediates (Bonilla *et al*, 2008). However, we recently performed single-cell RNA-sequencing (RNA-seq) of the developing midbrain and identified multiple cell types, including three different types of molecularly-diverse radial glia (Rgl1-3) with distinct temporal gene expression patterns and spatial distribution (La Manno *et al*, 2016). This finding has thus raised several important

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

questions such as: Which of these three radial glia cell types are capable of undergoing neurogenesis in the developing VM? Do they also contribute to neurogenesis with cell extrinsic factors? How are cell intrinsic and extrinsic factors spatially and temporally integrated? What is the cellular microenvironment or niche in which mDA neurogenesis takes place?

In order to address these questions, we used single cell and bulk RNA-sequencing as well as systems biology methodologies to gain further understanding of the cellular and molecular components of the mDA neurogenic niche. Our study identifies Rgl1-3 as the main cell types contributing signals and cells to mDA neurogenesis. However, each cell type contributed in a different way. While Rgl1 was identified as the neurogenic radial glia cell type, Rgl3 was the main contributor to cell extrinsic factors in the mDA neurogenic niche, including morphogens, growth factors and ECM. Our work thus provides the first unbiased and integrated analysis of the major molecular processes taking place in the mDA neurogenic niche. We also found how the expression of transcription factors, as well as signaling and ECM components, are controlled over time and in defined cell types forming the mDA neurogenic niche. We conclude that efforts aiming at recapitulating mDA neuron development in stem cells, either for cell replacement or for in vitro disease modeling, should thus focus on the generation of Rgl1 and Rgl3, the two main pillars of mDA neurogenesis.

RESULTS

Transcriptomic analysis of the embryonic ventral midbrain

In order to obtain a genome-wide view of the niche occupied by midbrain progenitor cells during mDA neurogenesis, we performed bulk RNA-seq of the mouse VM and surrounding tissues from embryonic day (E)11.5 to E14.5, the period at which mDA neurogenesis takes place (Arenas *et al*, 2015). We obtained transcriptomic profiles from five different regions of the neural tube of TH-GFP mouse embryos (Matsushita *et al*, 2002): the VM, the ventral hindbrain (HB), ventral forebrain (FB), dorsal midbrain (DM) and alar plate (L) (Fig. 1A). Specific anatomical and

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

temporal gene expression patterns were identified in all samples (Fig. 1B). A pair-wise correlation comparing all the VM samples revealed that E11.5 was the most divergent stage (Fig. 1C). Furthermore, principal component analysis (PCA) confirmed the similarity between E12.5 and E13.5 (Fig. 1D, Fig. EV1A). Multiple differentially expressed genes (DEG) were identified in the VM, compared to the other dissected brain regions (Table EV1). The VM transcriptome was further analyzed by weighted gene co-expression network analysis. Eight gene modules were identified and found to correlate with distinct expression patterns that describe the changes in transcriptome over time (Fig. EV1B-D, details in Table EV2). These modules can be summarized in three general developmental patterns. The first pattern (Fig. 1E, three modules in EV1B), consists of genes whose expression is increased over time, such as *Slc6a3* (DAT, dopamine transporter), a marker of mDA neuron maturation (Fauchey *et al*, 2000). Gene ontology (GO) analysis of this module showed enrichment in processes related to neuronal development and maturation, peaking at E14.5 (Fig. EV1F). The second pattern consists of genes whose expression peaks at middle time points E12.5-13.5 (Fig. 1F, two modules in EV1D), such as *Ncor2*, which represses the differentiation of neuronal precursors (Jepsen *et al*, 2007). GO analysis showed enrichment of genes involved in tissue growth and developmental processes (Fig. EV1E). Lastly, we identified a third pattern with decreasing gene expression during development, from E11.5 (Fig. 1G, three modules in EV1C). This included genes such as *Notch3* and *Hes5*, a ligand and an effector of the Notch signaling pathway, respectively (Louvi & Artavanis-Tsakonas, 2006). GO analysis showed enrichment in processes that describe cell proliferation (Fig. EV1G). These analyses thus provided both a detailed molecular insight and a general overview of the biological processes taking place in the developing VM during mDA neurogenesis, in which there is a shift from proliferation to progressive neuronal maturation and repression of progenitor pathways and functions.

The dopaminergic module: A network that defines the midbrain dopaminergic neuron niche

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To identify molecular processes that specifically occur during VM and mDA neurons, we performed a second unbiased weighted gene co-expression network analysis of the transcriptomes, but this time including both the VM and neighboring regions. This analysis generated 13 modules (or networks) of co-expressed genes, one being the "light green" module (Fig. EV2A) that had the most enriched DEG in the VM from E11.5 to E14.5 (Fig. EV2B and Table 1). Further analysis revealed that the top 5% of the interactions (1235 out of 24506) involved 181 genes (out of 374), which were enough to separate the VM samples at all the stages analyzed, as assessed by PCA (Fig. EV2C). This refined module contained many known genes expressed in the mDA lineage (*Foxa1*, *Shh*, *Wnt5a*, *Th*, *Ddc*, *Snca*, etc) and was thus designated as the mDA module. This module was represented as an undirected co-expression network (Fig. 2A), in which the color of the node indicates changes in gene expression (RPKM) over time and the lines linking them represent the pair-wise correlation of the linked genes (Langfelder & Horvath, 2008). GO analysis of the mDA module provided us with an overview of the biological processes in which these genes participate. These included relatively well-studied processes in mDA neuron development, such as dopamine biosynthetic process (p-value: 4.19e-9), morphogenesis of epithelium (p-value: 2.85e-5) or neural tube development (p-value: 1.91e-3); but also less well studied processes such extracellular matrix (p-value: 2.17e-7) or extracellular matrix component (p-value: 8.04e-3), amongst others (Fig. EV2D). These results indicate that the ECM is likely to play a more significant role than previously anticipated in mDA neuron development.

In order to identify the cell types contributing to the mDA module by gene expression, we used a single-cell RNA-seq analysis of the VM that recently allowed us to identify the twenty-six cell types present in the mouse VM during mDA neurogenesis (La Manno *et al*, 2016). This dataset was used to perform a cell type deconvolution of our mDA network, assigning cell type/s to the identified expression profiles in the mDA network (Fig. EV3A). 97% of the genes in the network (176 of 181 genes) were assigned to at least one cell type. Most of the genes (51.2%) were contributed by only four cell types present in the VZ: three types of radial glia like cells (Rgl 1-3)

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and ependymal cells (Fig. 2B, C). Sixteen additional cell types expressed the remaining 48.8% of the gene on the mDA network, while the remaining six other cell types did not contribute to this network. Notably, the four VZ cell types were the main contributors to the formation of the ECM (p-value: 7.72e-26) and its regulation (p-value: 2.88e-15), according to the gene sets defined in the Matrisome project (http://matrisomeproject.mit.edu/) (Naba *et al*, 2012). GO enrichment analysis confirmed that the main contribution of the four VZ cells types was in genes of ECM components, neural projections as well as in Shh and FGF signaling pathways (Fig. 2D). These results thus identify Rgl and ependymal cells as important contributors to the expression of ECM and critical signaling factors controlling mDA neuron development.

We next examined the specific contribution of the most abundant cell types in the mDA niche during neurogenesis, that is, radial glia, but not ependymal cells, which emerge later. Radial glia has been previously identified as a neurogenic cell in the mDA lineage (Bonilla *et al*, 2008) and as the source of multiple signaling molecules in the midbrain floor plate (Arenas *et al*, 2015). We thus decided to compare the transcriptome profiles of individual radial glia cell types, Rgl1-3, by gene set enrichment analysis (Subramanian *et al*, 2005) in order to explore whether distinct radial glia cell types may be poised to serve distinct functions. Analysis of Rgl1 revealed an enrichment in the expression of genes related to cell proliferation, such as M Phase (Normalized Enrichment Score (NES): 2.341, q-value <0.001), ribosomal constituents (NES: 2.641, q-value <0.001) and cellular biosynthetic process (NES: 1.992, q-value <0.001), which defined Rgl1 as a cell in a highly proliferative state. Rgl2, also expressed genes related to proliferation (Mitosis, NES: 1.759, q-value: 0.04), but additionally expressed genes involved in fatty acid metabolism (NES 1.837, q-value: 0.01) and cholesterol biosynthesis (NES: 1.893, q-value: 0.006), possibly linking Rgl2 to the production of specific midbrain cholesterol metabolites controlling mDA neurogenesis (Theofilopoulos *et al*, 2013). The last radial glia, Rgl3, was not enriched for proliferation genes, but rather in ECM core components (NES 2.029, q-value <0.001) and ECM regulators (NES 2.363, q-value <0.001) as well as signaling pathways (Glypican 1 Pathway, NES: 1.840, q-value: 0.02,

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Netrin 1 signaling, NES: 1.809, q-value: 0.04). These results suggested a high degree of specialization of radial glia and lead us to examine in further detail their contribution to the mDA niche with regards to secreted signaling molecules, such as morphogens and growth factors as well as ECM.

The extracellular matrix in the ventral midbrain

In order to further delve into the possible role of the ECM, we decided to identify the VM cell types responsible for the expression of genes coding for critical components of the ECM, regardless of their contribution to the mDA module. In order to rank the cell type contribution to the ECM, we developed a score based on the number of genes being expressed and their levels of expression (see methods), taking in account both core ECM components and ECM regulators, as defined by the Matrisome gene sets (Hynes & Naba, 2012; Naba *et al*, 2012). Six cell types (Rgl2-3, ependymal cells, endothelial, pericytes and microglia) were above a threshold set at 99.9% quantile of the mean ECM score (Fig. 3A). These cell types contributed to 82.8% of the total number of mRNA molecules of ECM core components (Fig. EV4A), and to 88.2% of ECM regulators in the VM (Fig. EV4B). Genes coding for core ECM components (Fig. 3B, EV4C) or ECM modifiers (Fig. 3C, EV4D) that were significantly expressed above baseline are shown in a color scale proportional to mRNA molecules for each cell type. Our results show that Rgl3, endothelial cells and pericytes were the main contributors to ECM core transcripts in the VM (Fig. 3A). While these 3 cell types shared some of the transcripts (*Sparc* and *Sparccl1*), each cell type expressed specific transcripts, such as *vitronectin, colagen4a1/2* and *agrin* in endothelia and pericytes or *netrin1, decorin* and *spondin1* in Rgl3 (Fig. 3B). The latter finding is particularly interesting as we identified Rgl3 as a key component of the mDA neurogenic niche (Fig. 2) and its products, Netrin1 and Decorin are known to regulate axonal development in mDA neurons (Kastenhuber *et al*, 2009) and midbrain neurogenesis (Long *et al*, 2016), respectively. On the other hand, we found that microglia is the main cell type involved in ECM regulation (Fig. 3D), in

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

particular, by expressing multiple types of cathepsin proteases (Fig. 3C), which likely contribute to remodel the ECM and allow them to migrate through the brain parenchyma (Chapman *et al*, 1997). Finally, from a temporal perspective, we found that Rgl1 and pericytes are more abundant during early neurogenesis (E11.5-12.5) while Rgl2-3 predominates during later stage (E14.5-E15.5) and ependymal cells subsequently after that at E18.5 (Fig. 3E). Adjusting the scores to the relative abundance of cell types at every stage confirmed the importance of the six cell types above and indicated an additional contribution of immature DA neurons (DA0) and Neuronal progenitors (NProg) at E11.5-E12.5 (Fig. S5 A-F). In summary, our results suggest a quantitatively important and constant contribution to the ECM by non midbrain-specific cell types, such as endothelial cells pericytes and microglia, suggesting a generic role of these cells. Instead, we found a midbrain- and stage-specific early contribution of NProg and Rgl1, and a late one by Rgl3.

Signaling pathways in the ventral midbrain

Our analysis of the mDA module in addition of suggesting an important contribution of radial glia to the ECM, also identified the cell types expressing signaling pathway components known to participate in mDA neuron development (Fig. 2B-D). To determine the possible directionality of the developmental signaling pathways, we curated a gene set of ligands and their corresponding receptors (see methods, Table EV3) for known signaling pathways and used that gene set to identify the cell types expressing ligands and their receptors in the VM. As in the case of the ECM, we examined the involvement of a cell type in signaling calculating a receptor score and a ligand score (see methods). This analysis identified seven cell types (Rgl1-3, ependymal cells, endothelial, pericytes and microglia) as the main contributors to signaling (Fig. 4A), covering multiple ligands (Fig. 4B) and receptor families (Fig. 4C), several of which are known for their role in embryonic midbrain development (Arenas *et al*, 2015; Smidt & Burbach, 2007). Notably, Rgl3 and pericytes showed the highest score for combined number and levels of ligands and receptors (Fig. 4A), which indicates that Rgl3 and pericytes are the main signaling centers in the VM. These

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cells offer the possibility of signaling in an autocrine/paracrine manner and of a cross-talk between the neural tissue and the vasculature, whose function and significance for mDA neuron development remains to be determined. On the other hand, Rgl1-2 expressed more receptors than ligands (Fig. 4A), which make them the main candidates to respond to and execute different aspects of VM and mDA neuron development.

As with the ECM score, we adjusted the signaling score to cell type abundance. This analysis suggests an early role of mRgl1 and progenitor cells at early stages (Fig. S5 G-I), and a late role of generic cell types such as endothelial cells and pericytes, as well as midbrain-specific cell types, such as mRgl2 and mRgl3 (Fig. S5 J-L).

Wnt signaling was represented with the highest number of types of ligands and receptors (Fig. 4B-C), underlining the prevalence of this pathway and its importance in mDA neuron development (Arenas, 2014). Interestingly, our analysis predicts novel candidate cell-to-cell signaling events with single-cell resolution (Fig. 4B-C). For instance, we found evidence for autocrine loops in multiple cell types, for nearly all pathways examined: Wnt, Notch, Hedgehog, BMP, VEGF, FGF, semaphorins and ephrin in Rgl1-3; CNTF in ependymal cells; TNF in microglia, IGF and PDGF in pericytes, and TGFβ in endothelial cells. No evidence for Slit-Robo and endothelin autocrine signaling was found, but rather for possible directional paracrine signaling from Rgl3 to endothelial cells (Slit-Robo) or from endothelial cells and microglia to Rgl1-3, ependymal and pericytes (endothelin) (Fig. 4B, C). Combined, our findings identify radial glia cells, in particular Rgl3, as the midbrain-specific cell type that expresses the most complete combination of ECM components and ligands for signaling pathways in midbrain development.

Transcriptional networks in Rgl2 and Rgl3.

We previously identified floor plate radial glial cells as mDA progenitors (Bonilla et al., 2008). However, our recent identification of three different types of radial glial cells in the ventral midbrain (La Manno et al., 2016) raised the question of which of the three identified radial glia cell

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

types is the mDA progenitor that undergoes mDA neurogenesis. We thus decided to examine the transcription factors expressed in each radial glia cell type using curated databases with identified target genes (Janky *et al*, 2014). Transcription factors in the database were filters to include only those expressed by Rgl1-3, and then we clustered them according to shared target genes (Fig. 5A, EV6A-B). We next investigated which of the possible transcription factor combinations can best explain the transcriptional profile of Rgl1-3 (Fig. 5B). Analysis of the statistical significance of the multiple combinations of transcription factors that up-regulate target genes in Rgl1-3 was performed using the Fast Westfall-Young (FWY) permutation test for multiple testing correction (Terada *et al*, 2013b). This procedure also prunes for non-significant combinations (Tarone, 1990), transforming the combinatorial analysis into a tractable problem. The specificity of the transcription factor combination analysis was confirmed by performing a control FWY analysis of transcription factors randomly selected from MSigDB while maintaining the transcriptional profile of each cell type. Random selections of transcription factors in Rgl1-3, resulted in much fewer significant combinations than obtained with the transcription factors expressed in each cell (Fig. EV6C). Thus, our results show that it is possible to obtain insights of the transcriptional networks responsible for the transcriptome of a cell type by integrating information about transcription factor binding sites and individual cell type transcriptomes.

Analysis of Rgl2 identified two large transcription factor clusters composed by Sox and proneural basic helix-loop-helix (bHLH) transcription factors (Fig. EV6A). We next investigated by which of the possible transcription factor combinations can best explain the transcriptional profile of Rgl2. However, FWY analysis showed that the only significant combination of transcription factors in Rgl2 was that formed by *PAX6* and *TCF7L1* (p-value=0.025), which has been reported to maintain neural progenitors in undifferentiated state (Kuwahara *et al*, 2014). Analysis of the target genes for these two transcription factors reveled enrichment of GO terms associated to development of the forebrain (p-value: 3.12e-45), hindbrain (p-value: 7.88e-8) and midbrain (p-value: 5.89e-5), suggesting a generic role in maintenance of neural progenitors. Moreover, Rgl2 cells and PAX6

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expression are found in the basal plate of the midbrain, a compartment that does not give rise to mDA neurons but is rather involved in glial cell development (La Manno *et al*, 2016).

Examination of Rgl3 also identified two large clusters formed by Sox and bHLH genes (Fig. EV6B). FWY analysis of Rgl3 revealed 15 enriched transcription factor combinations, which formed a network centered on *TEAD1*, a component of the hippo pathway (Harvey & Hariharan, 2012), as well as *RFX4, PDLIM5* and *SOX13* (Fig. EV6D-E). Analysis of the targets genes of these transcription factors, identified the IGF and Wnt signaling pathways, followed by the regulation ECM components (Fig. EV6F), which again underlines a role of Rgl3 in the formation of the mDA niche.

Identification of a transcriptional network in Rgl1 involved in mDA neurogenesis.

We next analyzed Rgl1, the radial glia that appears the earliest in the developing VM (Fig. 3E) as well as the one expressing less ECM (Fig. 3A) and more receptors than ligands for developmental pathways (Fig. 4A). As it was the case for Rgl1 and Rgl3, two large Sox and bHLH transcription factor clusters with common target genes were identified (Fig. 5A). However, our FWY analysis revealed that a total of 25 transcription factors generate 419 significant combinations (Table EV4). We used a network to represent the combinations of transcription factors. A pairwise interaction score was calculated for all significant transcription factor combinations, based on the frequency and p-value (Fig. 5C). The mean number of transcription factors per combination was 5.465 (Fig. 5D) and the five most common transcription factors were *ARNTL*, *E2F3*, *ASCL1*, *SOX5* and *TCF4* (Fig. 5E). The transcription factors E2F5 and E2F3, two cell cycle regulators (Chen *et al*, 2009; Trimarchi & Lees, 2002), showed the highest number of interaction partners (node degree, Fig. 5F), but these interactions were less frequent and with higher p-values. When node degrees were weighted for interaction score, *ARNTL* was identified as the central gene in the network (Fig. 5F), followed by *E2F3, ASCL1*, *E2F5, TCF3*, *TCF4*, *TCF12*, *SOX2, SOX5* and *SREBF1*. These results indicate that components of the bHLH neurogenic cluster, such as *ARNTL, ASCL1, TCF3,*

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TCF4 and TCF12, explain most aspects of the transcriptional state of Rgl1. Moreover, some of the genes in this cluster have been previously found to regulate mDA neurogenesis, such as *Ascl1,* which works in concert with *Neurog2* (Kele *et al*, 2006); or TCF3 and 4, that interact with active β-catenin to control mDA neurogenesis (Arenas, 2014); or *SREBF1,* a direct target of the nuclear receptors *Nr1h2* and *Nr1h3* (Schultz *et al*, 2000), that are required and sufficient for mDA neurogenesis (Sacchetti *et al*, 2009; Theofilopoulos *et al*, 2013).

Target genes of the transcription factors in this network were analyzed for enrichment using the curated gene sets of the molecular signature database repository (MSigDB, (Subramanian *et al*, 2005)). We found that the main functions controlled by this network were the regulation of cell cycle as well as Wnt and Notch signaling (Fig. EV6G), all of which are fundamental for mDA neurogenesis (Louvi & Artavanis-Tsakonas, 2006; Arenas *et al*, 2015). Analysis of putative target genes of this network included Wnt signaling components such as *DVL2*, *CTNNB1*, *GSK3B* and *TCF12* (Fig. 5G) (Inestrosa & Arenas, 2010; Willert & Jones, 2006; Neuman *et al*, 1993). *FOXA1,* involved in Shh signaling and mDA neuron development (Hynes *et al*, 1995), as well as components of other signaling pathways such as *IGF1R* (Quesada *et al*, 2007) and *BMP7* (Brederlau *et al*, 2002). Genes involved in neurogenesis, such as *NEUROD1* (Cho & Tsai, 2004), *HDAC2* (Jawerka *et al*, 2010) and *TCF12* (Uittenbogaard & Chiaramello, 2002). As well as the transcriptions factors required for mDA neuron development, such as *NFE2L1* (Villaescusa *et al*, 2016) and the nuclear receptor *NR4A2* (Zetterström *et al*, 1997), as well as markers of mature mDA neurons, *TH* and *ALDH1A1* (Arenas *et al*, 2015), were also found. Thus our results indicate that a transcriptional network mainly formed by *ARNTL, TCFs* (*3*, *4* and *12*), *ASCL1*, *SOXs* (*2* and *5*) and *SREBF1* can set in motion a transcriptional program allowing them to control cell cycle exit and neurogenesis, respond to Wnt and Shh signaling and differentiate into postmitotic neuroblasts and mDA neurons.

Combined, our analysis of transcription factor networks and their target genes suggest that each Rgl cell type contribute to diverse aspects of VM and mDA neuron development. While Rgl2

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expresses transcription factors known to control progenitor maintenance, Rgl3 expresses niche signals. Notably, the only radial glia cell type found to expresses a significant combination of transcription factors known to regulate mDA neurogenesis was Rgl1, suggesting that this is the cell type previously found to undergo mDA neurogenesis in the VM (Bonilla et al., 2008).

ARNTL is required for mDA neurogenesis

Our analysis of Rgl1 identified ARNTL as the central node and the most predominant transcription factor in a network that explains the transcriptional profile of Rgl1. *ARNTL* is known to promote neurogenesis, control cell cycle exit and regulate circadian rhythm (Bouchard-Cannon *et al*, 2013; Kimiwada *et al*, 2009; Malik *et al*, 2015). However, it is at present unknown whether it plays any role in VM development or mDA neurogenesis. We thus decided to examine whether *ARNTL* is both sufficient and required for mDA neuron development. For this purpose, we first examined the presence of ARNTL in the developing midbrain and found that it is present in $Sox2^+$ cells in the developing midbrain floor plate at E13.5 (Fig S7A). To test its functionality we then took advantage of a human long-term neuroepithelial stem (hLT-NES) cell line derived from human embryos (Sai2, Tailor *et al*, 2013), and performed *ARNTL* gain and loss of function experiments (Fig. 6A,B, S7B) during mDA neuron differentiation (Villaescusa *et al*, 2016). As previously reported, hLT-NES cells acquired a midbrain floor plate phenotype, assessed by the presence of both FOXA2 and LMX1A (Fig. 6C) and immunoreactivity for DA markers such as TH and NR4A2 at day 8 of differentiation (Fig. 6D).

For gain of function of experiments, proliferating hLT-NES cells were transduced at day -2 with a lentivirus carrying a tetracycline-inducible vector (Gossen *et al*, 1995) encoding for ARNTL, or GFP as control. Cells were treated with doxycycline from day-1 to day 1 and given a 4 h pulse of EdU before day 0, to mark proliferating cells that undergo mDA neurogenesis and become EdU^+ and TH^+ at day 8 after differentiation (Fig S7B). Analysis of ARNTL levels at day 0 revealed a 75% increase in protein levels compared to endogenous levels in GFP-expressing control cells (Fig.

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S7C). Notably, we found that *ARNTL* overexpression at this level did not affect the proportion of TH^+ cells undergoing neurogenesis (double EdU^+ and TH^+, Fig. S7D, E), but higher levels of ARNTL impaired their survival. These results suggested that ARNT overexpression "*per se*" is either not active in this process or that it cannot increase mDA neurogenesis beyond what endogenous levels ARNTL already achieve in Lt-NES cells. We thus examined whether endogenous ARNTL is required for mDA neurogenesis.

We next performed loss of function experiments using lentiviral vectors to stably knockdown endogenous ARNTL in hLT-NES cells. The resulting cell line, shARNTL-hLT-NES, exhibited a dramatic reduction of ARNTL protein levels, compared to a control shRNA (shControl, Fig. 6B). As with the gain of function, hLT-NES cells were pulsed with EdU before differentiation, to examine neurogenesis. Analysis of cells differentiated for 8 days revealed that the percentage of EdU^+ and TH^+ cells was reduced by 80%, from 14.23 ± 0.04% (mean ± SD) in shControl to 3.79 ± 0.017% in shARNTL-hLT-NES (Fig. 6E, F). Thus, our results show that endogenous ARNTL is *required* for DA neurogenesis and support the emerging concept that a network of bHLH transcription factors controls this process.

DISCUSSION

In this study, we have gained a broad view of the transcriptional programs operating in the VM and at single-cell level during mDA neurogenesis. Our analysis identifies a transcriptional program that defines mDA neuron development and Rgl1-3 as the main cell types expressing genes that regulate mDA neurogenesis and the mDA niche, providing thus novel insights into midbrain development at a single-cell level. Notably, we found that Rgl1 selectively expresses transcription factors controlling target genes required for mDA neurogenesis and cell membrane receptors for multiple developmental signals. On the other hand, Rgl2 was found to express transcription factors responsible progenitor maintenance, receptors for developmental factors and genes involved in lipid metabolism. Finally, Rgl3 was found to express transcription factors regulating the formation and

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

maintenance of the mDA niche, as well as multiple ECM components and ligands controlling several developmental pathways (Fig. 6). Thus, our study suggests that each of the three radial glia cell types (Rgl1-3) contributes to different aspects of mDA neuron development.

Our analysis of the VM transcriptome identified a mDA module that was mainly contributed to by seven cell types: Rgl1-3, ependymal cells, endothelial cells, pericytes and microglia. However, the most abundant cells contributing to mDA neurogenesis were Rgl1 (E11.5-E12.5), Rgl2 (E13.5-E14.5) and Rgl3 (E15.5). Moreover, two of them, Rgl1 and Rgl3, are present in the midbrain floor pate (La Manno *et al*, 2016), the anatomical compartment that generates mDA neurons (Arenas *et al*, 2015). While both cells contribute to the expression of developmental signals that control mDA neurogenesis, we found that Rgl3 is the main contributor to ligands and ECM. We also identify Rgl1 as the main cell type expressing receptors for midbrain developmental signals. Thus, our data suggest a model in which Rgl1 controls neurogenesis in an autocrine manner from E11.5 to E13.5 and Rgl3 in a paracrine fashion from E13.5-E14.5. This switch in neurogenic regulation may be of importance for mDA neuron development as *substantia nigra* mDA neurons emerge earlier than those in the ventral tegmental area. We thus examined the expression of ligands in early radial glia (Rgl1) versus late radial glia (Rgl3) for pathways whose receptors are expressed in Rgl1. While no change in expression was detected for ligands such as *Shh, Sema3b, Sema5b, Efna4* and *Efnb1*, a switch was detected from *Wnt7a and Wnt7b* in Rgl1 to *Wnt5a and Wnt5b* in Rgl3. In addition, other factors were expressed in early Rgl1, such as *Jag1*, or in late Rgl3, such as *Sfrp1, Dkk3, Bmp1, Fgf7, Ntn1, Dcn* and *Spon1*. These results suggest that one of the major changes taking place in early vs late mDA neuron niche is a switch form early Wnt/β-catenin signaling by Wnt7a (Fernando *et al*, 2014) to late Wnt/β-catenin-independent signaling by Wnt5a (Andersson *et al*, 2008). This switch was further confirmed by the late action of two pro-differentiation Wnt/β-catenin inhibitors such as Sfrp1 (Kele *et al*, 2012) and Dkk3 (Fukusumi *et al*, 2015), reinforcing thus a role for Wnt/β-catenin-independent signaling at late stages. The necessity of an appropriate balance between these two pathways has been recently confirmed by analysis of double *Wnt1* and

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wnt5a mutant mice (Andersson *et al*, 2013). Our study thus extends to Wnt5b and Wnt7b the spectra of candidate Wnt ligands participating in this complex balance and identifies the precise cell types expressing Wnts and their receptors during mDA neuron development.

We also found that cell types that are not midbrain specific, such as endothelial cells, pericytes and microglia, express factors relevant for mDA neurogenesis, such as *Wnt5b* and *Jag1* (Fig. 4A). Notably, the vasculature is known to be in contact with radial glia cells and their process during development, allowing thus for direct or indirect modulation of mDA neurogenesis. We thus suggest that these cells can be part of a generic neurogenic niche, that may add on to midbrain-specific signals derived from radial glia.

Our bioinformatics analysis of the transcription factor networks in Rgl1-3 also suggest individual functions for each radial glia cell type. These range from controlling neurogenesis (Rgl1), to progenitor maintenance (Rgl2) and niche formation (Rgl3). Notably, our results suggest that the radial glia previously identified to undergo mDA neurogenesis (Bonilla *et al*, 2008), is Rgl1. Indeed, Rgl1 expressed a neurogenic bHLH network formed by *Arntl, Ascl1*, *Tcf3*, *4*, *12* and *Srebf1*. Previous studies have linked transcription factors such as *Ascl1* to mDA neurogenesis, via *Neurog2* (Kele *et al*, 2006), or *Tcf3/4* regulated by Wnt/β-catenin signaling (Arenas, 2014) or *Srebf1* (Schultz *et al*, 2000) by *Nr1h2-3/Lxr* (Sacchetti *et al*, 2009; Theofilopoulos *et al*, 2013). Moreover, the Rgl1 network also includes two members of the Sox family of transcription factors. *Sox2*, which is expressed in floor plate progenitors (Kele *et al*, 2006; Bonilla *et al*, 2008) and *Sox5*, which promotes cell cycle exit and differentiation (Martinez-Morales *et al*, 2010). Notably, this network was centered on the clock gene *ARNTL*, also known as *Bmal1*, a pioneer transcription factor (Menet *et al*, 2014). ARNTL controls cell cycle entry/exit and neurogenesis (Bouchard-Cannon *et al*, 2013; Malik *et al*, 2015) and directly regulates the expression of *Neurod1* (Kimiwada *et al*, 2009), a proneural transcription factor that initiates neuronal differentiation and migration (Peyton *et al*, 1996; Pataskar *et al*, 2015). *ARNTL* had not been previously linked to mDA neurogenesis, but based on key role predicted by our network analysis, we investigated the role of this transcription factor

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and found it is required to promote mDA neurogenesis. Thus, our results identify ARNTL as a novel factor controlling mDA neuron development, and support the idea that several bHLH transcription factors control mDA neurogenesis.

In sum, our study examined the transcriptome of the developing VM during mDA neurogenesis and identifies two radial glia cell types, Rgl1 and Rgl3, as the main components of the mDA neurogenic niche. While Rgl1 expresses transcription factors and target genes required for mDA neurogenesis, Rgl3 expresses core ECM components and most ligands required to control midbrain development and mDA neurogenesis. These cells are part of an extended niche formed by a neural component (Rgl1-3 and ependymal cells), a non-neural component (endothelial cells and pericytes and microglia). Our results thus uncover the diversity and richness of the molecular components of the mDA niche, their cellular origin and their temporal dynamics during mDA neurogenesis. Moreover, our results identify novel components of the mDA niche, including a novel gene required for mDA neurogenesis, *ARNTL*, and provide new knowledge that can be applied to improve current regenerative medicine approaches for the treatment of Parkinson's disease.

AUTHOR CONTRIBUTIONS

EMT performed the biological validation experiments. EMT and GLM performed the data analysis. GLM and SI performed the RNA extraction and cDNA library preparation. GLM ran the NGS pipeline. DG and CV performed the embryonic dissections. PRdVC performed the preparation of lentiviral particles. SL helped to design the experiments and interpreted results. EMT and EA designed experiments, interpreted results and wrote the manuscript. All authors have read and approved the manuscript.

ACKNOWLEDGMENTS

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We thank members of the Arenas lab for their help and suggestions; and Johnny Söderlund and Alessandra Nanni for technical and secretarial assistance. Knut and Alice Wallenberg Foundation for support to the CLICK imaging facility at KI. Financial support was obtained from Swedish Research Council (VR projects: DBRM, 2008:2811, 2011-3116 and 2011-3318), Swedish Foundation for Strategic Research (SRL program), European Commission (NeuroStemcellRepair and DDPD-Genes), Karolinska Institutet (SFO Thematic Center in Stem cells and Regenerative Medicine) and Hjärnfonden to EA, grants from the Swedish Research Council and EU - FP7 program DDPDGENES to S.L (www.ddpdgenes.eu) and a fellowship from the Swedish Research Council to E.M.T.

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

Bulk transcriptome determination. Mouse embryos were obtained from TH-GFP animals (Matsushita *et al*, 2002) that were mated overnight, and noon of the day the plug was considered E0.5. Embryos were dissected out of the uterine horns at E11.5-E14.5 and placed in ice-cold sterile PBS where brain regions were dissected under a stereomicroscope with a UV attachment to detect GFP. VM samples corresponded to domains M3 to M7 of the floor and basal plate (Nakatani *et al*, 2007). Tissue samples were collected in separate tubes, stored at -80˚C until RNA isolation. Ethical approval for mice experimentation was granted by the local ethics committee, Stockholm Norra Djurförsöksetiska Nämnd number N326/12 and N158/15.

Total RNA isolation was performed with RNeasy kit (Qiagen). The RNA integrity and concentration was checked using Qubit and 2200 TapeStation (Agilent). Illumina TruSeq libraries were prepared using kit and protocols from Illumina. High-throughput sequencing was performed on a HiSeq 2000.

Genes Expression Analysis. Differentially expressed genes (DEG) on each developmental stage were identified utilizing Qlucore Omics Explorer v3.1 (Qlucore AB, Lund, Sweden) utilizing t-test comparing VM samples versus other regions, with false discovery rate q-value correction (Storey, 2002). Variance filter of 15% and a threshold for significance of p-value = 0.01 across stages, unless otherwise stated. Sample correlation was calculated from log2(RPKM+1) values after variance filtering of 12.5%. Single-cell expression data was obtained from La Manno et al. 2016. Significant expression is baseline (>99.8% posterior probability). Analyses were made using R (https://www.r-project.org) and ggplot2 for plots (http://ggplot2.org/).

Weighted Gene Co-Expression Network Analysis (WGCNA) (Langfelder & Horvath, 2008) was performed with the log2(RPKM+1) transformed values, filtered by variance until 10,068 genes were selected. The topological overlap matrix was calculated with the variables of a signed network, with power of 7. The identification of modules was performed with the “tree” option on a minimum module size of 100; modules with correlation higher than 90% were merged. Module

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enrichment for DEG in VM at all analyzed stages was done with Fisher's exact test with false discovery rate q-value correction (Storey, 2002). A score for the enrichment was calculated as the product between the –log10(q-value) and a standardized z-score for DEG per module. Module network layout was made using Cytoscape and interactions in the top 5% of adjacency were selected for further analysis. Expression changes over time are represented by the color of each node, which was calculated as the normalized difference of RPKM between late (E13.5, E14.5) and early (E11.5, E12.5) for each gene.

The identification of VM gene modules with developmental or stage dependent expression were identified with WGCNA on the VM samples from E11.5 to E14.5. Further filtering for genes not expressed in the VM was performed until 9,061 genes were selected. The topological overlap matrix was calculated with the options of a signed network, with power of 17. The identification of modules was performed with the "tree" option on a minimum module size of 100 genes, modules with correlation higher than 99% were merged. Correlation with samples trait and Student asymptotic p-values were calculated as described (Langfelder & Horvath, 2008). Embryonic stage (E11.5 to E14.5) was used as sample trait, ordinal values for stage (1 for E11.5, 2 for E12.5 and so on), and binary values for middle stages (E12.5, E13.5 as 1). Network layouts and analysis, were made using Cytoscape v3.3.0 (Shannon *et al*, 2003) or Gephi v0.9.1 (Bastian *et al*, 2009). WGCNA were carried out with the R package v1.49 (Langfelder & Horvath, 2008).

Score for cell types contribution to ECM or signaling. Cell type contribution for ECM or signaling were calculated with the data from single cell of mouse VM (La Manno *et al*, 2016). The identified transcripts corresponded to cells found in the floor plate and basal plate of the midbrain, domains M3 to M7 (Nakatani *et al*, 2007). We selected the two components for each analysis. ECM was the result of the ECM core genes and ECM regulation gene sets (Naba *et al*, 2012). For signaling, curated ligands and receptors gene sets (Table EV3).

The score for a gene set is calculated as:

$$S_{gset,ct,Stage} = \alpha_{ct,Stage} * \frac{\sum_{g \in gset} M_{g,ct}}{\sum_{c \in ctypes} \sum_{g \in set} M_{g,c}} + * \alpha_{ct,Stage} * \frac{\sum_{g \in gset} N_{g,ct}}{\sum_{c \in ctypes} \sum_{g \in set} N_{g,c}} \quad \textbf{(1)}$$

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Let $M_{g,ct}$ be the Bayesian estimate of the expression level of the gene *g* in cell type *ct* and let $N_{g,ct}$ be the value of the indicator vector that is 1 when the gene *g* is expressed in cell type *ct* above baseline and 0 otherwise. Let α_{ct} be the relative sampling ratio of cell type *ct* in a developmental stage, with a value of 1 when the analysis include all the developmental stages. The score is an indicator between diversity of genes in a biological function or gene set and the expression levels of those genes. A combined score by the sum of both components give us an ECM or signaling scores. This simplification of a biological process in a tissue, provide an indicator for a biological processes break down by the cell types found in that tissue. The threshold line was set at the percentile 99.9% of the bootstrapped distribution of the mean of the combined score with 10e5 replicates.

Pathway receptor ligand. A list of ligand (159 genes) and receptors or co-receptor (114 genes) was curated from the literature, representing 21 signaling pathways. Ligands and receptors were grouped in pathways, without assuming specific ligand-receptor pair (Table EV3).

Network single-cell deconvolution. Using data from La Manno et al., each gene in the network was assigned to the cell types that expressed that gene, as identified above baseline (see above and (La Manno *et al*, 2016)). Cell types represented by one gene, or cell types with a final contribution less than 1% of the genes were excluded from the network.

Single-Cell Transcription Factor pattern mining. Transcription factors expressed on each radial glia cells were used as input to obtain the target of the human homologous from iRegulon plugin for Cytoscape (Janky *et al*, 2014), retrieving up to 1,000 target genes per transcription factor. Targets of transcription factors by cell type detailed on table EV5. Clustering of the transcription factor was done by Jaccard index, as an indicator of shared target genes.

Analysis of enrichment of combinatorial transcription factor target genes were done using Fast Westfall-Young (FWY) permutation procedure v1.0.1 (Terada *et al*, 2013a, 2013b). This is a computational efficient procedure for multiple testing correction with higher detection sensitivity than commonly used algorithms (Terada *et al*, 2013b). Upregulated or enriched genes among Rgl cells types was scored using the logarithmic difference between the groups (Subramanian *et al*,

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2005). Values above 0.5 (or 50% of upregulation) were considered enriched in that cell type. Targets of the transcription factors were tested for enrichment in the upregulated genes in the one-tailed Fisher test with a significance threshold of 0.05, and tested with 1,000 permutations in the FWY to generate a null distribution of randomly permutated datasets (Terada *et al*, 2013b).

The resulting significant combinatorial patterns were represented as a network on which the edge between a set of transcription factors is quantified by an interaction score. This interaction score is calculated for each transcription factor pair as the sum of –log10 (adjusted p-value) for all the transcription factor combinations of that pair. Adjusted p-values <0.001 were consider equal to 0.001 for calculations of interaction score. Controls were performed with randomly selected transcription factors from MSigDB C3 database (Subramanian *et al*, 2005). Selecting the same numbers of transcription factors used before for Rgl1-3 respectively. For each cell type, 100 random transcription factors combinations were analyzed by FWY maintaining all other setting.

Gene enrichment. Single-cell Gene Set Enrichment Analysis (GSEA) was done ranking the gene by the class difference for the cell type of interest. The analysis was performed on the GSEA software v2.2.2 (Subramanian *et al*, 2005) with the MSigDB genesets v5.0 for canonical pathways and GO biological processes (Subramanian *et al*, 2005). Due to the nature of the single-cell transcriptome profiles and molecular counting with unique molecular identifiers (Islam *et al*, 2014), negative enrichment score (ES) on a gene set for a cell type in particular have no meaning. Instead, negative ES were interpreted as enrichment on another compared cell type. Analysis of Gene Ontology (GO) enrichment were made utilizing the R package ClusterProfiler (Yu *et al*, 2012) or MSigDB (Subramanian *et al*, 2005) with false discovery rate q-value correction (Storey, 2002). Enrichment of the targets of transcription factor (Fig. EV6CF) were analyzed with hypergeometric test and with false discovery rate q-value correction (Storey, 2002) over the MSigDB C2 gene sets (Subramanian *et al*, 2005).

Human neuroepithelial stem cell differentiation. Sai2 hLT-NES cells were maintained as described of hLT-NES cell (Tailor *et al*, 2013), DA differentiation and lentivirus infection were

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

performed as previously described (Villaescusa *et al*, 2016). Each experimental condition was examined in three separate experiments and duplicate determinations.

Commercially available ARNTL shRNA lentiviral particles targeted against the human transcript were used (sc-38165, Santa Cruz Biotech.) to generate stable cell lines selected with puromycin (500 ng/ml first 4 days, maintained with 200 ng/ml). As negative control, lentiviral particles against no know human mRNA were used (sc-108080, Santa Cruz Biotech.). Expression levels of ARNTL were analyzed by western blot of total cell lysates, with anti-Arntl (1:1000, ab93806, Abcam). For loading control, anti-Lamin-B1 was used (1:5000, ab16048, Abcam). Differentiation experiments were done with stable shRNA cells with less than 10 passages counted from the lentiviral infection.

Mouse Arntl cDNA-pcDNA3.1 expression vector (Etchegaray *et al*, 2003) (kindly donate by Professors Steven Reppert and David Weaver, University of Massachusetts Medical School), was cloned into the lentiviral backbone Tet-O-FUW-EGFP (Addgene #30130) by blunt-end ligation into EcoRI blunted site of the backbone, replacing the EGFP open reading frame. The final plasmid was verified by sequencing before usage. Lentiviral production, infection and immunofluorescence were performed as previously described (Villaescusa *et al*, 2016; di Val Cervo *et al*, 2017). Inducible expression of ARNTL was done with doxycycline (Sigma, 50 µg/ml for day -1 and day 0, and 25 µg/ml for day 1).

Immunofluorescence was captured with a confocal microscope Zeiss LSM700 with a 10X 0.45NA objective, at a resolution of 0.3126 µm/pixel. A minimum of nine images per well were captured, with two technical replicates per condition in independent biological experiments (N=3 for LOF experiments, and N=4 for GOF experiments). Quantifications of EdU and DAPI positive nuclei were done utilizing Cell Profiler v2.2.0 (Jones *et al*, 2008). Double positive cell for TH and EdU, and total TH cells were counted manually in a condition-blind manner. This was done by randomizing each file name with *blindanalysis* perl script (Salter, 2016). Full images had linear levels adjusted for better visualization, done in Fiji (Schindelin *et al*, 2012).

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

Andersson ER, Prakash N, Cajanek L, Minina E, Bryja V, Bryjova L, Yamaguchi TP, Hall AC, Wurst W & Arenas E (2008) Wnt5a regulates ventral midbrain morphogenesis and the development of A9-A10 dopaminergic cells in vivo. *PLoS One* **3:** e3517

Andersson ER, Saltó C, Villaescusa JC, Cajanek L, Yang S, Bryjova L, Nagy II, Vainio SJ, Ramirez C, Bryja V & Arenas E (2013) Wnt5a cooperates with canonical Wnts to generate midbrain dopaminergic neurons in vivo and in stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **110:** E602-10

Arenas E (2014) Wnt signaling in midbrain dopaminergic neuron development and regenerative medicine for Parkinson's disease. *J. Mol. Cell Biol.* **6:** 42–53

Arenas E, Denham M & Villaescusa JC (2015) How to make a midbrain dopaminergic neuron. *Development* **142:** 1918–1936

Bastian M, Heymann S & Mathieu J (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM* **8:** 361–362

Bonilla S, Hall AC, Pinto L, Attardo A, Götz M, Huttner WB & Arenas E (2008) Identification of midbrain floor plate radial glia-like cells as dopaminergic progenitors. *Glia* **56:** 809–20

Bouchard-Cannon P, Mendoza-Viveros L, Yuen A, Kærn M & Cheng H-YM (2013) The circadian molecular clock regulates adult hippocampal neurogenesis by controlling the timing of cell-cycle entry and exit. *Cell Rep.* **5:** 961–73

Brederlau A, Faigle R, Kaplan P, Odin P & Funa K (2002) Bone morphogenetic proteins but not growth differentiation factors induce dopaminergic differentiation in mesencephalic precursors. *Mol. Cell. Neurosci.* **21:** 367–78

Chapman H a, Riese RJ & Shi GP (1997) Emerging roles for cysteine proteases in human biology. *Annu. Rev. Physiol.* **59:** 63–88

Chen H-Z, Tsai S-Y & Leone G (2009) Emerging roles of E2Fs in cancer: an exit from cell cycle control. *Nat. Rev. Cancer* **9:** 785–797

Cho J-H & Tsai M-J (2004) The role of BETA2/NeuroD1 in the development of the nervous system. *Mol. Neurobiol.* **30:** 35–47

Etchegaray J, Lee C, Wade P a & Reppert SM (2003) Rhythmic histone acetylation underlies transcription in the mammalian circadian clock. *Nature* **421:** 177–182

Fauchey V, Jaber M, Bloch B & Le Moine C (2000) Dopamine control of striatal gene expression during development: relevance to knockout mice for the dopamine transporter. *Eur. J. Neurosci.* **12:** 3415–25

Fernando C, Kele J, Bye C, Niclis JC, Alsanie W, Blakely BD, Stenman J, Turner B & Parish C (2014) Diverse roles for Wnt7a in ventral midbrain neurogenesis and dopaminergic axon morphogenesis. *Stem Cells Dev.* **23:** 1991–2003

Fukusumi Y, Meier F, Gotz S, Matheus F, Irmler M, Beckervordersandforth R, Faus-Kessler T, Minina E, Rauser B, Zhang J, Arenas E, Andersson E, Niehrs C, Beckers J, Simeone A, Wurst W & Prakash N (2015) Dickkopf 3 Promotes the Differentiation of a Rostrolateral Midbrain Dopaminergic Neuronal Subset In Vivo and from Pluripotent Stem Cells In Vitro in the Mouse. *J. Neurosci.* **35:** 13385–13401

Gossen M, Freundlieb S, Bender G, Müller G, Hillen W & Bujard H (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* **268:** 1766–9

Götz M & Huttner WB (2005) The cell biology of neurogenesis. *Nat Rev Mol Cell Biol* **6:** 777–788

Harvey KF & Hariharan IK (2012) The Hippo pathway. *Cold Spring Harb. Perspect. Biol.* **4:**

Hynes M, Porter JA, Chiang C, Chang D, Tessier-Lavigne M, Beachy PA & Rosenthal A (1995) Induction of midbrain dopaminergic neurons by sonic hedgehog. *Neuron* **15:** 35–44

Hynes RO & Naba A (2012) Overview of the matrisome-An inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* **4:**

Inestrosa NC & Arenas E (2010) Emerging roles of Wnts in the adult nervous system. *Nat. Rev. Neurosci.* **11:** 77–86

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P & Linnarsson S (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11:** 163–6

Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, Marine J-C & Aerts S (2014) iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput. Biol.* **10:** e1003731

Jawerka M, Colak D, Dimou L, Spiller C, Lagger S, Montgomery RL, Olson EN, Wurst W, Göttlicher M & Götz M (2010) The specific role of histone deacetylase 2 in adult neurogenesis. *Neuron Glia Biol.* **6:** 93–107

Jepsen K, Solum D, Zhou T, McEvilly RJ, Kim H-J, Glass CK, Hermanson O & Rosenfeld MG (2007) SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron. *Nature* **450:** 415–9

Jones DL & Wagers AJ (2008) No place like home: anatomy and function of the stem cell niche. *Nat. Rev. Mol. Cell Biol.* **9:** 11–21

Jones TR, Kang IH, Wheeler DB, Lindquist RA, Papallo A, Sabatini DM, Golland P & Carpenter AE (2008) CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics* **9:** 482

Kastenhuber E, Kern U, Bonkowsky JL, Chien C-B, Driever W & Schweitzer J (2009) Netrin-DCC, Robo-Slit, and heparan sulfate proteoglycans coordinate lateral positioning of longitudinal dopaminergic diencephalospinal axons. *J. Neurosci.* **29:** 8914–26

Kele J, Andersson ER, Villaescusa JC, Cajanek L, Parish CL, Bonilla S, Toledo EM, Bryja V, Rubin JS, Shimono A & Arenas E (2012) SFRP1 and SFRP2 dose-dependently regulate midbrain dopamine neuron development in vivo and in embryonic stem cells. *Stem Cells* **30:** 865–75

Kele J, Simplicio N, Ferri ALM, Mira H, Guillemot F, Arenas E & Ang S-L (2006) Neurogenin 2 is required for the development of ventral midbrain dopaminergic neurons. *Development* **133:** 495–505

Kimiwada T, Sakurai M, Ohashi H, Aoki S, Tominaga T & Wada K (2009) Clock genes regulate neurogenic transcription factors, including NeuroD1, and the neuronal differentiation of adult neural stem/progenitor cells. *Neurochem. Int.* **54:** 277–285

Kuwahara A, Sakai H, Xu Y, Itoh Y, Hirabayashi Y & Gotoh Y (2014) Tcf3 represses Wnt-β-catenin signaling and maintains neural stem cell population during neocortical development. *PLoS One* **9:** e94408

Langfelder P & Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9:** 559

Long K, Moss L, Laursen L, Boulter L & Ffrench-Constant C (2016) Integrin signalling regulates the expansion of neuroepithelial progenitors and neurogenesis via Wnt7a and Decorin. *Nat. Commun.* **7:** 10354

Louvi A & Artavanis-Tsakonas S (2006) Notch signalling in vertebrate neural development. *Nat. Rev. Neurosci.* **7:** 93–102

Lu P, Takai K, Weaver VM & Werb Z (2011) Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb. Perspect. Biol.* **3:** a005058–a005058

Malik A, Kondratov R V., Jamasbi RJ & Geusz ME (2015) Circadian clock genes are essential for normal adult neurogenesis, differentiation, and fate determination. *PLoS One* **10:** 1–20

La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, Borm LE, Stott SRW, Toledo EM, Villaescusa JC, Lönnerberg P, Ryge J, Barker RA, Arenas E & Linnarsson S (2016) Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167:** 566–580.e19

Martinez-Morales PL, Quiroga AC, Barbas J a & Morales A V (2010) SOX5 controls cell cycle progression in neural progenitors by interfering with the WNT-beta-catenin pathway. *EMBO Rep.* **11:** 466–472

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Matsushita N, Okada H, Yasoshima Y, Takahashi K, Kiuchi K & Kobayashi K (2002) Dynamics of tyrosine hydroxylase promoter activity during midbrain dopaminergic neuron development. *J. Neurochem.* **82:** 295–304

Menet JS, Pescatore S & Rosbash M (2014) CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev.* **28:** 8–13

Naba A, Clauser KR, Hoersch S, Liu H, Carr SA & Hynes RO (2012) The Matrisome: In Silico Definition and In Vivo Characterization by Proteomics of Normal and Tumor Extracellular Matrices. *Mol. Cell. Proteomics* **11:** M111.014647-M111.014647

Nakatani T, Minaki Y, Kumai M & Ono Y (2007) Helt determines GABAergic over glutamatergic neuronal fate by repressing Ngn genes in the developing mesencephalon. *Development* **134:** 2783–93

Neuman T, Keen A, Knapik E, Shain D, Ross M, Nornes HO & Zuber MX (1993) ME1 and GE1: basic helix-loop-helix transcription factors expressed at high levels in the developing nervous system and in morphogenetically active regions. *Eur J Neurosci* **5:** 311–318

Pataskar A, Jung J, Smialowski P, Noack F, Calegari F, Straub T & Tiwari VK (2015) NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.***:** 1–22

Peyton M, Stellrecht CM, Naya FJ, Huang HP, Samora PJ & Tsai MJ (1996) BETA3, a novel helix-loop-helix protein, can act as a negative regulator of BETA2 and MyoD-responsive genes. *Mol. Cell. Biol.* **16:** 626–33

Quesada A, Romeo HE & Micevych P (2007) Distribution and localization patterns of estrogen receptor-beta and insulin-like growth factor-1 receptors in neurons and glial cells of the female rat substantia nigra: localization of ERbeta and IGF-1R in substantia nigra. *J. Comp. Neurol.* **503:** 198–208

Sacchetti P, Sousa KM, Hall AC, Liste I, Steffensen KR, Theofilopoulos S, Parish CL, Hazenberg C, Richter LA, Hovatta O, Gustafsson J-A & Arenas E (2009) Liver X receptors and oxysterols promote ventral midbrain neurogenesis in vivo and in human embryonic stem cells. *Cell Stem Cell* **5:** 409–19

Salter J (2016) blindanalysis: v1.0.

Scadden DT (2006) The stem-cell niche as an entity of action. *Nature* **441:** 1075–9

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P & Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9:** 676–82

Schultz JR, Tu H, Luk A, Repa JJ, Medina JC, Li L, Schwendner S, Wang S, Thoolen M, Mangelsdorf DJ, Lustig KD & Shan B (2000) Role of LXRs in control of lipogenesis. *Genes Dev.* **14:** 2831–2838

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B & Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13:** 2498–504

Smidt MP & Burbach JPH (2007) How to make a mesodiencephalic dopaminergic neuron. *Nat. Rev. Neurosci.* **8:** 21–32

Storey JD (2002) A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **64:** 479–498

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES & Mesirov JP (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102:** 15545–15550

Tailor J, Kittappa R, Leto K, Gates M, Borel M, Paulsen O, Spitzer S, Karadottir RT, Rossi F, Falk A & Smith A (2013) Stem cells expanded from the human embryonic hindbrain stably retain regional specification and high neurogenic potency. *J Neurosci* **33:** 12407–12422

Tarone RE (1990) A modified Bonferroni method for discrete data. *Biometrics* **46:** 515–22

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Taverna E, Götz M & Huttner WB (2014) The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu. Rev. Cell Dev. Biol.* **30:** 465–502

Terada A, Okada-Hatakeyama M, Tsuda K & Sese J (2013a) Statistical significance of combinatorial regulations. *Proc. Natl. Acad. Sci. U. S. A.* **110:** 12996–3001

Terada A, Tsuda K & Sese J (2013b) Fast Westfall-Young permutation procedure for combinatorial regulation discovery. *Proc. - 2013 IEEE Int. Conf. Bioinforma. Biomed. IEEE BIBM 2013***:** 153–158

Theofilopoulos S, Wang Y, Kitambi SS, Sacchetti P, Sousa KM, Bodin K, Kirk J, Saltó C, Gustafsson M, Toledo EM, Karu K, Gustafsson J-Å, Steffensen KR, Ernfors P, Sjövall J, Griffiths WJ & Arenas E (2013) Brain endogenous liver X receptor ligands selectively promote midbrain neurogenesis. *Nat. Chem. Biol.* **9:** 126–33

Trimarchi JM & Lees JA (2002) Sibling rivalry in the E2F family. *Nat. Rev. Mol. Cell Biol.* **3:** 11–20

Uittenbogaard M & Chiaramello a (2002) Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. *Brain Res. Gene Expr. Patterns* **1:** 115–121

di Val Cervo PR, Romanov RA, Spigolon G, Masini D, Martín-Montañez E, Toledo EM, La Manno G, Feyder M, Pifl C, Ng Y, Sánchez SP, Linnarsson S, Wernig M, Harkany T, Fisone G & Arenas E (2017) Induction of functional dopamine neurons from human astrocytes in vitro and mouse astrocytes in a Parkinson's disease model. *Nat. Biotechnol.*

Villaescusa JC, Li B, Toledo EM, Rivetti di Val Cervo P, Yang S, Stott SR, Kaiser K, Islam S, Gyllborg D, Laguna-Goya R, Landreh M, Lönnerberg P, Falk A, Bergman T, Barker RA, Linnarsson S, Selleri L & Arenas E (2016) A PBX1 transcriptional network controls dopaminergic neuron development and is impaired in Parkinson's disease. *EMBO J.* **35:** 1963–1978

Willert K & Jones KA (2006) Wnt signaling: Is the party in the nucleus? *Genes Dev.* **20:** 1394–1404

Yu G, Wang L-G, Han Y & He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16:** 284–7

Zetterström RH, Solomin L, Jansson L, Hoffer BJ, Olson L & Perlmann T (1997) Dopamine neuron agenesis in Nurr1-deficient mice. *Science* **276:** 248–50

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

Figure 1. Transcriptomic profile of the ventral midbrain.

(**A**) Embryonic tissue dissection scheme and regions collected for RNA-seq from TH-GFP embryos of stages E11.5, E12.5, E13.5 and E14.5. VM, ventral midbrain. DM, dorsal midbrain. FB, forebrain floor plate. HB, hindbrain floor plate. L, alar plate. (**B**) Heat map representation of high variance genes (rows) in dataset of all samples (columns) across all time points. Samples are color coded for stage and region. (**C**) Correlation of VM samples after filtering for variance (12.5%). (**D**) First principal component analysis of VM samples across all time points. Full plot is in Fig. EV1A. (E-G) Gene expression patterns in the developing VM as obtained by WGCNA of the VM samples. (**E**) Pattern summarizing gene modules with positive correlation over development (top left) and examples of genes: *Slc6a3*, *Ostbp2* and *Gria1*. (**F**) Pattern summarizing gene modules correlating at E12.5 and E13.5 and examples of genes in these modules: *Lrp1*, *Ncor2* and *6230400d17Rik*. (**G**) Pattern summarizing gene modules with negative correlation over development and some examples: *Hes5*, *Notch3* and *Otx2*. Expression is in RPKM. Red lines represent the median expression of each module. All modules with significant correlations are shown in Fig. EV1B-D.

Figure 2. The dopaminergic module and its deconvolution at a single-cell level.

(**A**) Weighted gene co-expression network analysis of the mDA module, filtered for the top 5% of interactions. Color represents changes in levels of gene expression during development. Node size is proportional to the mean expression levels of the gene during development. (**B**) Contribution of the different cell types in the VM to the network formed by the dopaminergic module. (**C**) Diagram of VM floor plate with focus on the cell types present in the VZ. (**D**) GO enrichment terms corresponding to genes with the highest contribution to the mDA module. Top, GO for biological process. Bottom, GO for cellular components.

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3. Analysis of the contribution of individual mouse VM cell types to the composition and dynamics of the extracellular matrix.

(**A**) Cell type contribution to the ECM as determined by scores for ECM core components and ECM regulators. Dotted line represents quantile 99.9% of the bootstrapped mean of the ECM scores. (**B**) Heat map of the expression of core components of the ECM by the most significant cell types. Color intensity is proportional to the Bayesian estimate of expression level. Gray scale indicates values below significance level. mRgl1-3, radial glia type 1 to 3; mPeri, pericyte; mEndo, endothelial cell; mEpend, ependymal cell; mMgl, microglia. (**C**) Heat map of the expression of genes regulating the ECM regulator genes by the most significant cell types, as described as above. Color boxes under each gene identify their ontology. (**B**, **C**) The horizontal plots represent the total number of molecules per cell type. (**D**) Percentage of genes expressed by each cell type that participate in the regulation of the ECM (green) versus core ECM components (purple). (**E**) Relative abundance of cell types at different developmental stages, as sampled by single cell RNA-sequencing of the VM (La Manno *et al*, 2016).

Figure 4. Contribution of individual mouse VM cell types to signaling in the mDA niche.

(**A**) Plot showing the receptor and ligand scores of the different VM cell types. Color is proportional to the signaling score. Dotted line represents quantile 99.9% of the bootstrapped mean of the signaling scores. (**B**) Heat map representation of the expression of receptors in identified VM cell types. (**C**) Heat map representation of the ligands expressed by the most cell types in the VM. Gene colored according to their activity; green for activation, red for inhibition and yellow for context-dependent modulation. (**B, C**) Color intensity is proportional to the Bayesian estimate of expression level. Gray scale indicates values below significance level. mRgl1-3, radial glia type 1 to 3; mPeri, pericyte; mEndo, endothelial cell; mEpend, ependymal cell; mMgl, microglia.

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5. Combinatorial analysis of the enrichment of transcription factors in Rgl1.

(**A**) Clustering of transcription factors expressed in Rgl1 by Jaccard index of shared target genes. A bar plot to the right represents the number of targets genes for each transcription factor. (**B**) Diagram of the combinatorial analysis of transcription factors (TF) by their enrichment in common targets genes. Left, enriched genes are those regulated by TF1-4 and expressed above threshold. Red lines represent upregulated target genes. Several target genes were shared by different TF. Right, network representation of TF sharing target genes, analyzed by Fast Westfall-Young (FWY) multiple combinatorial transcription factor enrichment. (**C**) Network representation of FWY analysis result of Rgl1. Node color is proportional to node degree and its size is proportional to weighted degree. Nodes with high weighted degrees (core nodes) have a black border. Color intensity and width of the line connecting two nodes (edge) is proportional to the interaction score of the TF pair. (**D**) Distribution of number of TF involved in significant combinations. (**E**) Frequency distribution of significant TF permutations distributed accordingly to their abundance. (**F**) Plot comparing node degree and weighted node degree of the network. For presentation purpose, weighted degree is shown in logarithmic scale. Color intensity and size is proportional to weighted degree. (**G**) Heat map of selected target genes of core transcription factors enriched on Rgl1 (red) found in databases.

Figure 6. ARNTL is required for the differentiation of human neuroepithelial stem cell into mDA neurons. (**A**) Representation of the protocol to examine dopaminergic neurogenesis in hLT-NES cells. (**B**) Western blot analysis identified the presence of ARNTL in the mouse ventral midbrain at E11.5 (left) and in control hLT-NES cells. ARNTL was dramatically reduced by shRNAs against ARNTL (right). LAMIN-B1 was used as loading control. (**C**) The midbrain floor plate marker FOXA2 (green) and the DA lineage marker LMX1A (red) are present in hLT-NES at day 8 of differentiation. (**D**) hLT-NES differentiated for 8 days are immunoreactive for the DA markers TH (green) and NR4A2 (red). (**E**) Analysis of neurogenesis in shControl and shARNT

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

hLT-NES cells as identified by the presence of cells double positive (arrows) for TH (green cytoplasm) and EdU (red nuclei). (**F**) Quantification of mDA neurogenesis: TH^+ and EdU^+ cells relative to the total TH+ cells. (p-value = 0.0362, N = 3). Scale bars in C, D and E, 100 µm.

EXTENDED VIEW FIGURE LEGENDS

Figure EV1. Ventral midbrain developmental genes modules.

(**A**) Principal component analysis of VM samples at different developmental times. Insert shows the percentage of variance by component. (**B**) Modules with positive correlation of gene expression with development. (**C**) Modules with negative correlation to development. (**D**) Modules that correlate at E12.5 and E13.5. Correlations and p-values are detailed in each plot. (**E-G**) GO analysis of genes modules correlating at stages E12.5 and E13.5 (**E**), or with positive (**F**) or negative (**G**) correlation during development.

Figure EV2. Weighted Gene Co-Expression Network Analysis of all samples.

(**A**) Dendrogram of modules obtained from weighted gene co-expression network analysis. The mDA module is shown in “light green” and has black border. (**B**) Heat map of q-value of DEG in the mDA module for each developmental stage. Red scale, q-values between 0 and 0.05; black to light gray scale, q-values above 0.05. (**C**) Principal component analysis of the genes corresponding to the mDA module “light green”. The genes in the module were sufficient to cluster VM samples from the rest of the tissue samples. Color represents tissue and size represents developmental stage. (**D**) GO analysis of genes in the mDA module: Top, biological processes; Bottom, cellular components.

Figure EV3. Single-cell network deconvolution.

(**A**) Identity matrix of VM cell types and genes expressed in the mDA module. Blue lines represent significantly expressed genes in the corresponding cell type, compared to baseline. Red, gene

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

increased; Yellow, maintained; Green, decreased over time. Color scale in figure 2a. (**B**) Analysis of gene expression levels (RPKM) in the mDA module for each cell type, comparing early (E11.5+E12.5) versus late stages (E13.5+E14.5).

Figure EV4. Contribution of different ventral midbrain cell types to the expression of extracellular matrix components and modifiers.

(**A-B**) Bar plot with the contribution of each cell type to the total number of transcripts for core ECM genes (**A**) and regulatory ECM genes (**B**). (**C**) Heat map of ECM core component genes by cell type. Color intensity is proportional to the Bayesian estimate of expression level. Gray scale indicates values below significance level. Bar plot to the right represents total average of transcripts for ECM core components per cell type. (**D**) Heat map of the genes regulating ECM composition per cell type, represented as described above.

Abbreviations: mMgl, microglia. mRgl1-3, radial glia type 1 to 3. mEndo, endothelial. mPeri, pericyte. mEpend, ependymal. mGaba1a, 1b and 2, GABAergic neurons 1a, 1b and 2. mSert, serotonergic neurons. mNbL1-2, lateral neuroblasts. mNbML1-5 mediolateral neuroblasts. mRN, Red Nucleus. mOMTN oculomotor and trochlear neurons. mNProg, neuronal progenitor. mNbM, medial neuroblast. mNbDA DA neuroblast. mDA0-2, DA neurons 1-2.

Figure EV5. Stage dependent contribution of individual mouse VM cell types to signaling and ECM score in the mDA niche.

(**A-F**) Plots showing the receptor and ligand scores of the different VM cell types considering cell type abundance at each developmental stages studied. (**G-L**) Plots showing the ECM regulator and ECM core score considering cell type abundance for each developmental stage. (**A, G**) E11.5, (**B, H**) E12.5, (**C, I**) E13.5, (**D, J**) E14.5, (**E, K**) E15.5 and (**F, L**) E18.5. Color intensity is proportional to the signaling score. Dotted line at mean plus standard deviation of the mean of the scores.

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure EV6. Combinatorial enrichment of transcription factors on Rgl2 and Rgl3.

(**A**) Clustering of transcription factors expressed in Rgl2 by Jaccard index of shared target genes. (**B**) Clustering of transcription factors expressed in Rgl3 by Jaccard index of shared target genes. (**C**) Percentage of FWY significant combinations obtained with random selection of transcription factors compared to FWY results of Rgl1-3 transcription profiles. (**D**) Network representation of FWY analysis of Rgl3. Node color and size are proportional to node degree. Nodes with higher weighted degrees (core nodes) have a black border. Color intensity and width of the lines connecting the nodes are proportional to the interaction score for each transcription factor pair. (**E**) Plot comparing node degree and weighted node degree of the network obtained by FWY analysis of Rgl3. Color intensity and size are proportional to the weighted degree. (**F**) Gene enrichment of target genes for core transcription factors in Rgl3. Analysis was performed with MSigDB gene set C2 canonical pathways v5.0. (**G**) Gene enrichment of target genes for core transcription factors in Rgl1. Analysis was done with MSigDB gene set C2 canonical pathways v5.0.

Figure EV7. Overexpression of ARNTL in human neuroepithelial stem cells during dopaminergic differentiation.

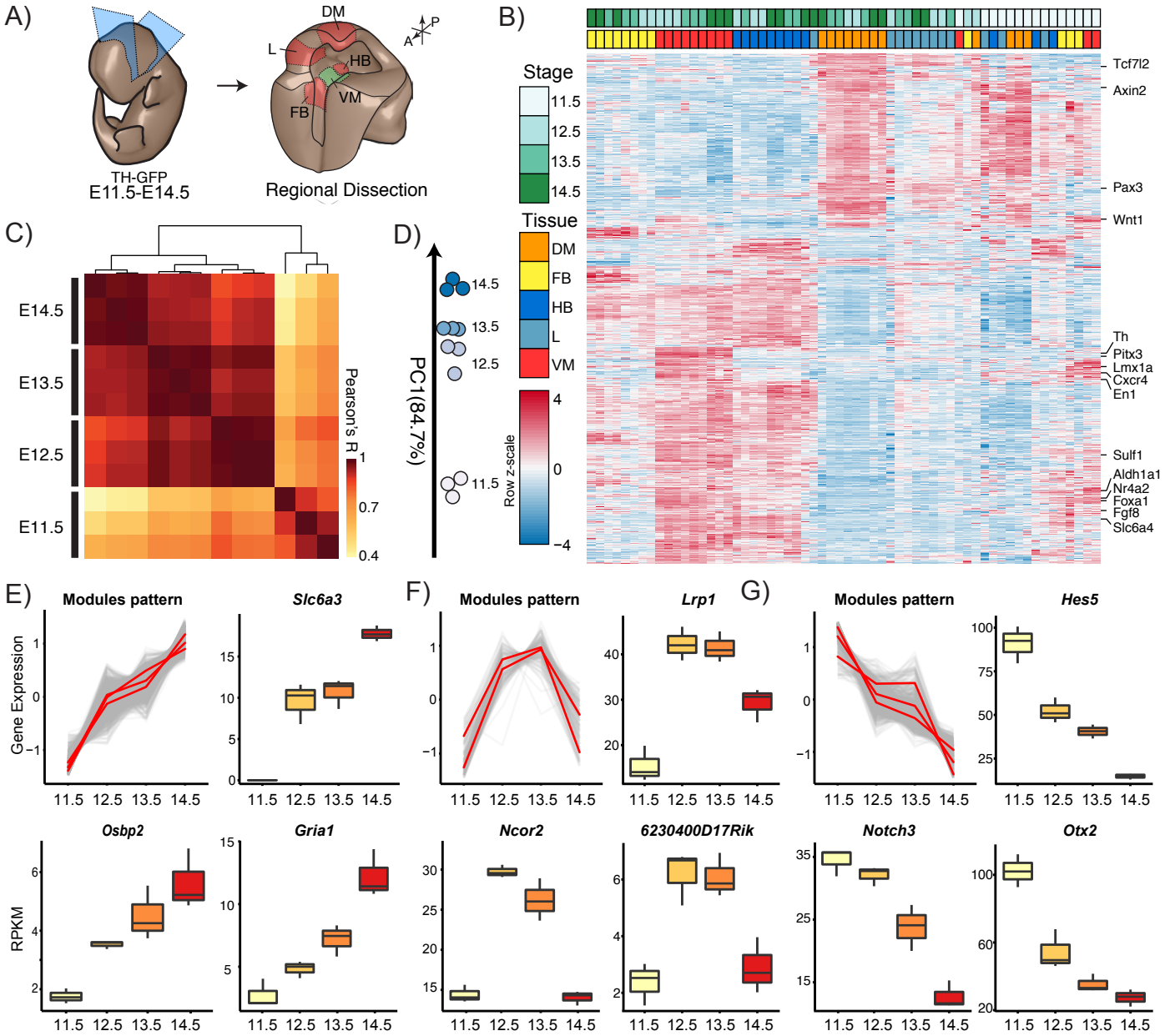
(**A**) ARNTL protein (green) is present in SOX2+ cells (red) in the developing mouse VM at E13.5. DAPI in blue. Scale bar, 100 µm. Image to the right magnified from box to the left. Full arrow heads, double positive SOX2 and ARNTL cells. Empty arrowheads, SOX2 and ARNTL negative cells. (**B**) Protocol for the overexpression of ARNTL in hLT-NES cells. (**C**) Increased levels of ARNTL were detected 24h after doxycycline treatment to LT-NES cells infected with ARNTL-lentivirus, compared to EGFP as control hLT-NES cells. LAMIN-b1 was used as loading control. (**D**) Representative immunofluorescence images corresponding to control and ARNT transduced hLT-NES cells at day 8. Cells having undergone mDA neurogenesis (arrowhead) are identified by the incorporation of EdU (red) in TH^+ cells (green). Scale bar, 50 µm. (**E**) Quantification of mDA neurogenesis: TH^+ and EdU^+ cells relative to the total TH^+ cells (n.s, non-significant, N = 4).

bioRxiv preprint doi: https://doi.org/10.1101/155846; this version posted June 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLE

Table 1

Module	Number VM DEG	Module size	q-value	Enrichment Score
Black	37	551	9.12e-02	-0.55
Brown	46	856	1.08e-01	-0.14
Cyan	6	281	9.15e-04	-5.68
Green	119	2839	9.26e-03	6.10
Green Yellow	10	423	1.80e-04	-6.35
Grey	24	519	3.99e-02	-1.53
Grey60	55	175	1.03e-24	5.80
Light cyan	18	201	3.26e-02	-2.01
Light green	191	378	5.88e-136	824.97
Light yellow	2	100	2.46e-02	-3.28
Magenta	69	3087	3.56e-17	13.90
Salmon	54	287	6.10e-14	2.63
Tan	11	371	2.63e-03	-4.27


A)
TH-GFP
E11.5-E14.5
DM
L
HB
VM
FB
A
P
Regional Dissection
B)
Stage
11.5
12.5
13.5
14.5
Tissue
DM
FB
HB
L
VM
Row z-scale
4
0
−4
Tcf7l2
Axin2
Pax3
Wnt1
Th
Pitx3
Lmx1a
Cxcr4
En1
Sulf1
Aldh1a1
Nr4a2
Foxa1
Fgf8
Slc6a4
C)
E14.5
E13.5
E12.5
E11.5
Pearson's R
1
0.7
0.4
D)
PC1(84.7%)
14.5
13.5
12.5
11.5
E)
Modules pattern
Gene Expression
Slc6a3
Osbp2
Gria1
RPKM
F)
Modules pattern
Lrp1
Ncor2
6230400D17Rik
G)
Modules pattern
Hes5
Notch3
Otx2
11.5 12.5 13.5 14.5

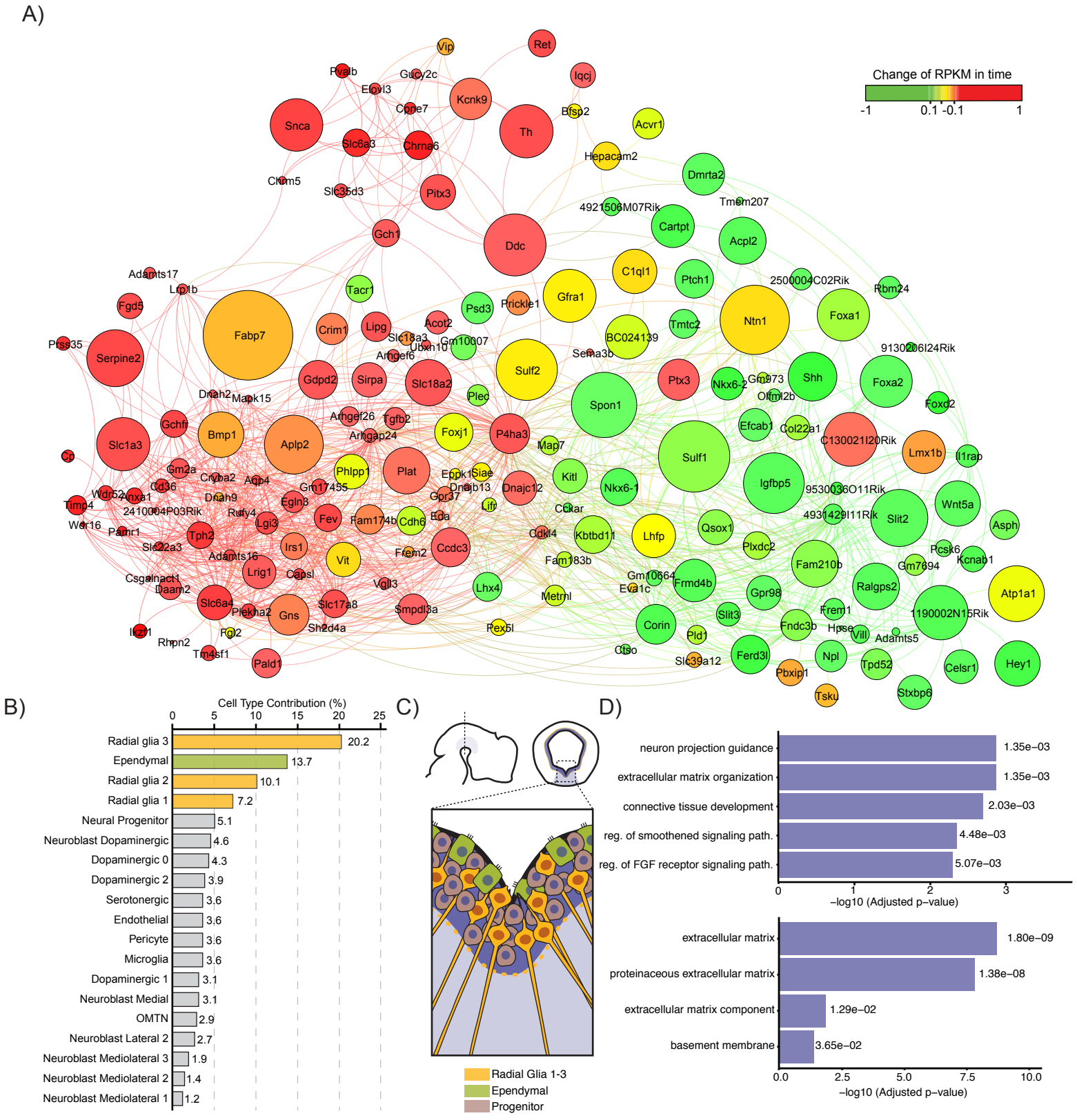

A)
Change of RPKM in time
-1
0.1 -0.1
1
Vip
Ret
Pvalb
Gucy2c
Elovl3
Kcnk9
Iqcj
Cpne7
Bfsp2
Snca
Th
Acvr1
Slc6a3
Chrna6
Hepacam2
Chrm5
Slc35d3
Pitx3
Dmrta2
4921506M07Rik
Tmem207
Cartpt
Gch1
Ddc
Acpl2
C1ql1
Ptch1
Adamts17
Lrp1b
Tacr1
Gfra1
2500004C02Rik
Rbm24
Fgd5
Psd3
Prickle1
Foxa1
Fabp7
Crim1
Lipg
Acot2
Ntn1
Tmtc2
Prss35
Serpine2
Slc18a3
Gm10007
BC024139
Ubxn10
Arhgef6
Sema3b
9130206I24Rik
Sulf2
Gdpd2
Sirpa
Slc18a2
Ptx3
Nkx6-2
Gm973
Shh
Foxa2
Dnah2
Mapk15
Plec
Spon1
Olfml2b
Foxd2
Arhgef26
Tgfb2
Gchfr
Bmp1
Foxj1
P4ha3
Efcab1
Col22a1
Slc1a3
Aplp2
Arhgap24
Map7
C130021I20Rik
Lmx1b
Il1rap
Cp
Sulf1
Gm2a
Phlpp1
Plat
Eppk1
Siae
Kitl
Igfbp5
Cryba2
Aqp4
Gm17455
Dnajb13
Dnajc12
Nkx6-1
9530036O11Rik
Wdr52
Cd36
Anxa1
Dnah9
Gpr37
Timp4
Egln3
Lifr
Wnt5a
2410004P03Rik
Rufy4
Cckar
4931429I11Rik
Slit2
Wdr16
Lgi3
Fev
Fam174b
Cdh6
Eda
Asph
Pamr1
Tph2
Cdkl4
Kbtbd11
Lhfp
Qsox1
Slc22a3
Irs1
Ccdc3
Plxdc2
Pcsk6
Adamts16
Frem2
Fam183b
Fam210b
Kcnab1
Vit
Gm7694
Lrig1
Capsl
Gm10664
Csgalnact1
Vgll3
Lhx4
Frmd4b
Ralgps2
Atp1a1
Daam2
Eva1c
Gpr98
Slc6a4
Metrnl
Slc17a8
Frem1
Plekha2
Gns
Smpdl3a
Slit3
1190002N15Rik
Corin
Fndc3b
Hpse
Ikzf1
Sh2d4a
Pex5l
Pld1
Vill
Adamts5
Fgl2
Rhpn2
Ctso
Ferd3l
Npl
Tm4sf1
Slc39a12
Tpd52
Celsr1
Hey1
Pald1
Pbxip1
Tsku
Stxbp6
B)
Cell Type Contribution (%)
0
5
10
15
20
25
Radial glia 3 20.2
Ependymal 13.7
Radial glia 2 10.1
Radial glia 1 7.2
Neural Progenitor 5.1
Neuroblast Dopaminergic 4.6
Dopaminergic 0 4.3
Dopaminergic 2 3.9
Serotonergic 3.6
Endothelial 3.6
Pericyte 3.6
Microglia 3.6
Dopaminergic 1 3.1
Neuroblast Medial 3.1
OMTN 2.9
Neuroblast Lateral 2 2.7
Neuroblast Mediolateral 3 1.9
Neuroblast Mediolateral 2 1.4
Neuroblast Mediolateral 1 1.2
C)
Radial Glia 1-3
Ependymal
Progenitor
D)
neuron projection guidance 1.35e−03
extracellular matrix organization 1.35e−03
connective tissue development 2.03e−03
reg. of smoothened signaling path. 4.48e−03
reg. of FGF receptor signaling path. 5.07e−03
0
1
2
3
−log10 (Adjusted p−value)
extracellular matrix 1.80e−09
proteinaceous extracellular matrix 1.38e−08
extracellular matrix component 1.29e−02
basement membrane 3.65e−02
0.0
2.5
5.0
7.5
10.0
−log10 (Adjusted p−value)

Figure 2

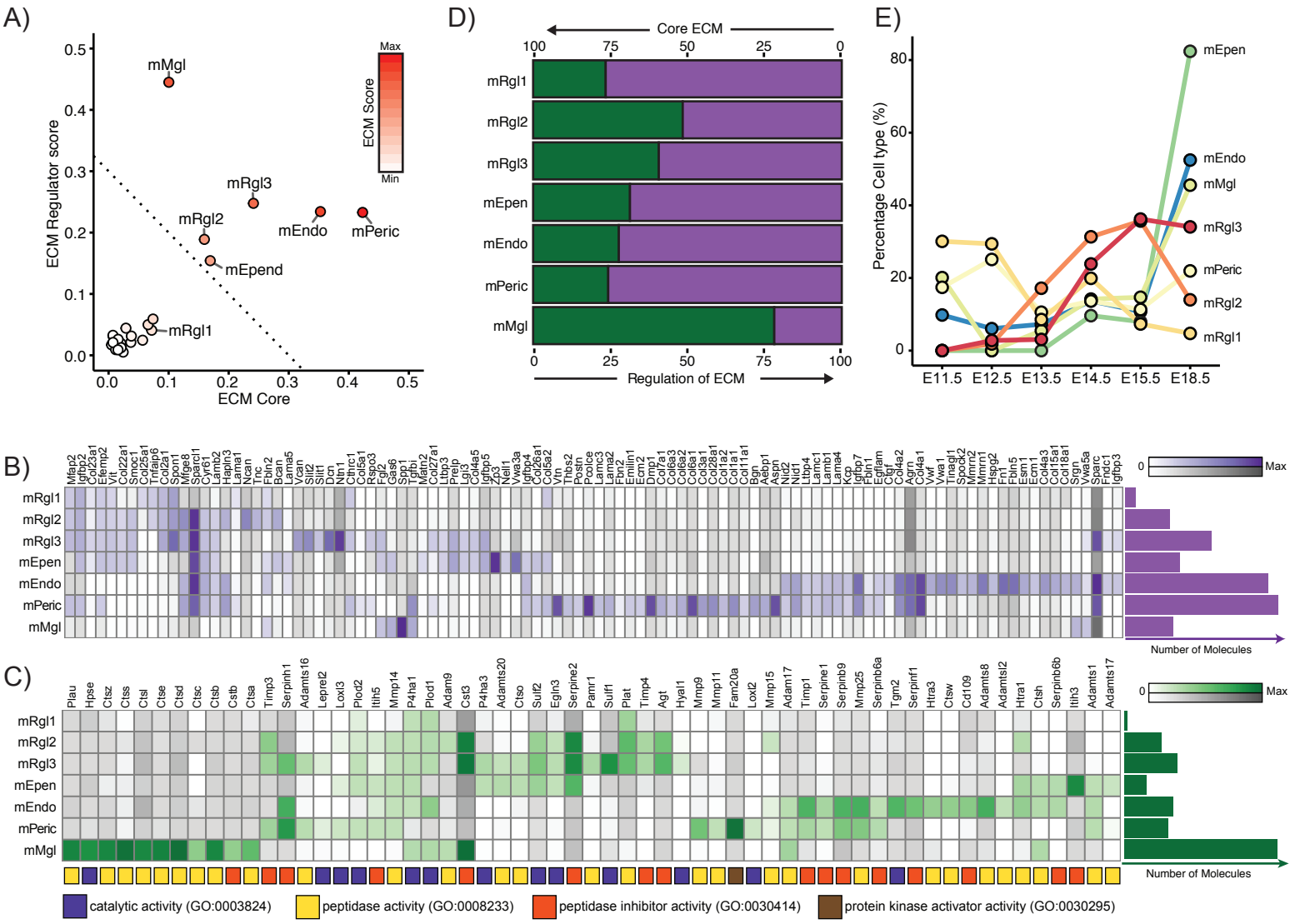

A)
ECM Regulator score
ECM Core
ECM Score
Max
Min
mMgl
mRgl3
mRgl2
mEndo
mPeric
mEpend
mRgl1
D)
Core ECM
Regulation of ECM
mRgl1
mRgl2
mRgl3
mEpen
mEndo
mPeric
mMgl
E)
Percentage Cell type (%)
E11.5
E12.5
E13.5
E14.5
E15.5
E18.5
mEpen
mEndo
mMgl
mRgl3
mPeric
mRgl2
mRgl1
B)
Max
Number of Molecules
C)
Max
Number of Molecules
catalytic activity (GO:0003824)
peptidase activity (GO:0008233)
peptidase inhibitor activity (GO:0030414)
protein kinase activator activity (GO:0030295)

Figure 3

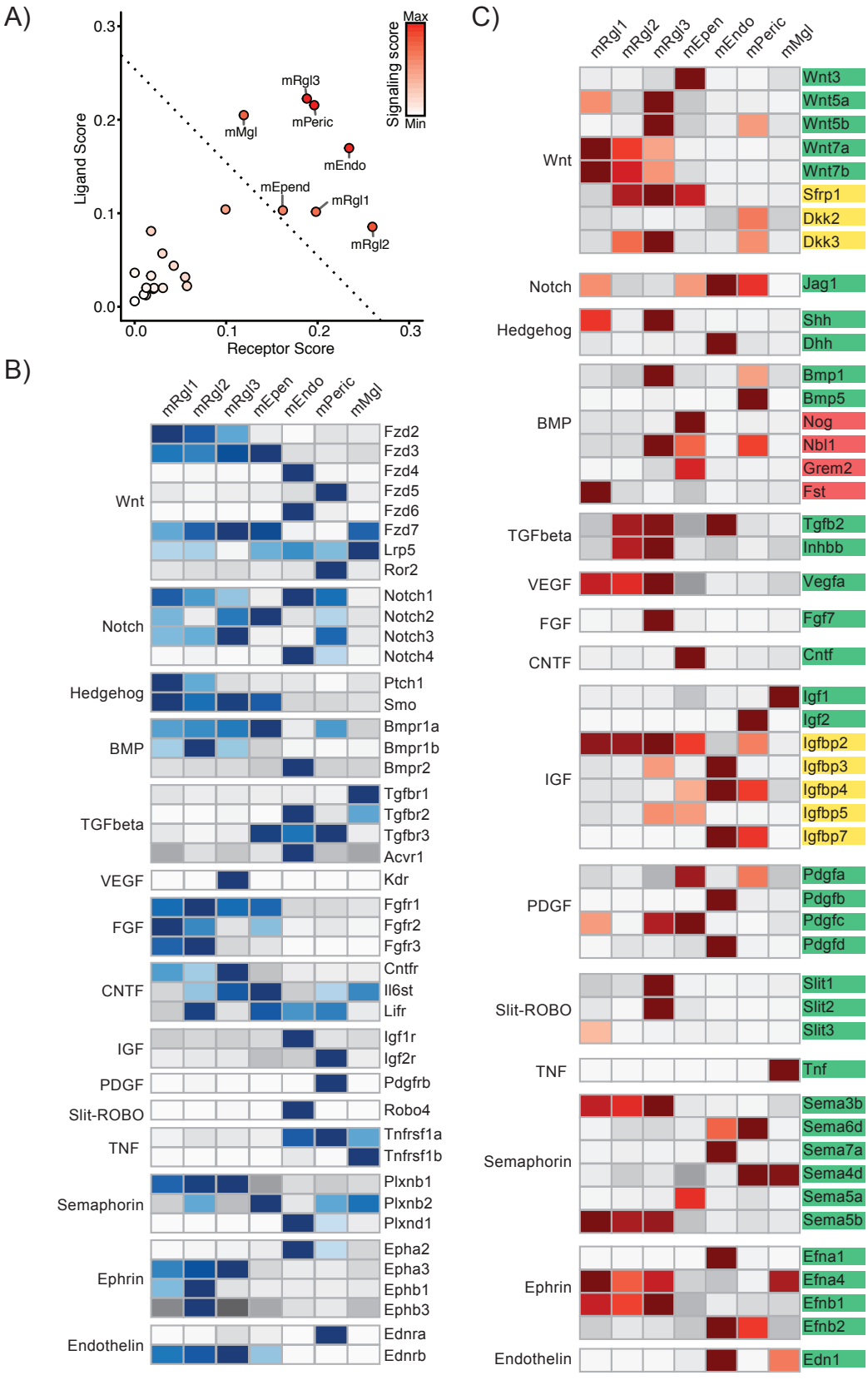

A)
Ligand Score
Receptor Score
Signaling score
Max
Min
mRgl3
mPeric
mMgl
mEndo
mEpend
mRgl1
mRgl2
B)
mRgl1
mRgl2
mRgl3
mEpen
mEndo
mPeric
mMgl
Wnt
Fzd2
Fzd3
Fzd4
Fzd5
Fzd6
Fzd7
Lrp5
Ror2
Notch
Notch1
Notch2
Notch3
Notch4
Hedgehog
Ptch1
Smo
BMP
Bmpr1a
Bmpr1b
Bmpr2
TGFbeta
Tgfbr1
Tgfbr2
Tgfbr3
Acvr1
VEGF
Kdr
FGF
Fgfr1
Fgfr2
Fgfr3
CNTF
Cntfr
Il6st
Lifr
IGF
Igf1r
Igf2r
PDGF
Pdgfrb
Slit-ROBO
Robo4
TNF
Tnfrsf1a
Tnfrsf1b
Semaphorin
Plxnb1
Plxnb2
Plxnd1
Ephrin
Epha2
Epha3
Ephb1
Ephb3
Endothelin
Ednra
Ednrb
C)
mRgl1
mRgl2
mRgl3
mEpen
mEndo
mPeric
mMgl
Wnt
Wnt3
Wnt5a
Wnt5b
Wnt7a
Wnt7b
Sfrp1
Dkk2
Dkk3
Notch
Jag1
Hedgehog
Shh
Dhh
BMP
Bmp1
Bmp5
Nog
Nbl1
Grem2
Fst
TGFbeta
Tgfb2
Inhbb
VEGF
Vegfa
FGF
Fgf7
CNTF
Cntf
IGF
Igf1
Igf2
Igfbp2
Igfbp3
Igfbp4
Igfbp5
Igfbp7
PDGF
Pdgfa
Pdgfb
Pdgfc
Pdgfd
Slit-ROBO
Slit1
Slit2
Slit3
TNF
Tnf
Semaphorin
Sema3b
Sema6d
Sema7a
Sema4d
Sema5a
Sema5b
Ephrin
Efna1
Efna4
Efnb1
Efnb2
Endothelin
Edn1

Figure 4

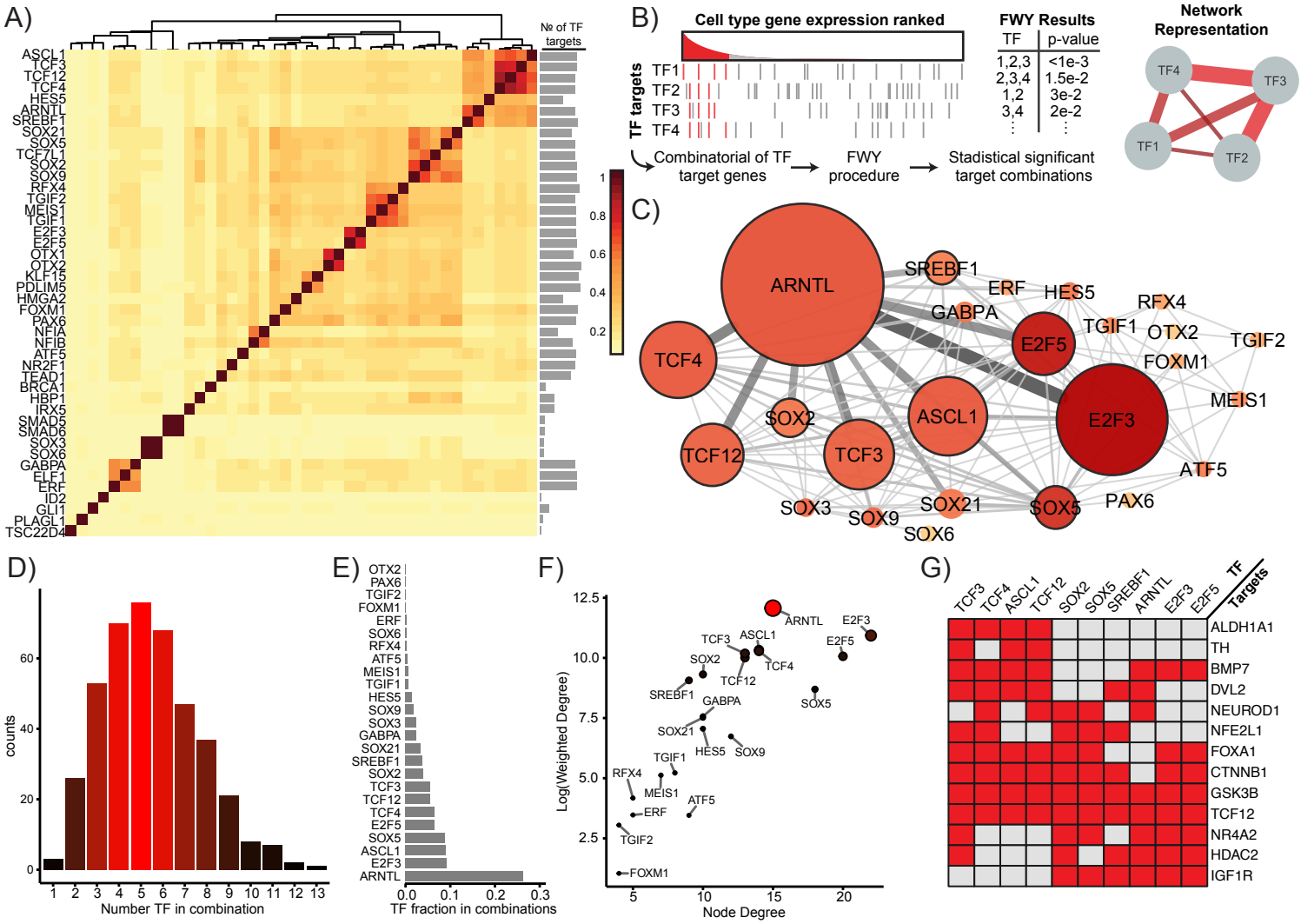

A)
№ of TF targets
ASCL1
TCF3
TCF12
TCF4
HES5
ARNTL
SREBF1
SOX21
SOX5
TCF7L1
SOX2
SOX9
RFX4
TGIF2
MEIS1
TGIF1
E2F3
E2F5
OTX1
OTX2
KLF15
PDLIM5
HMGA2
FOXM1
PAX6
NFIA
NFIB
ATF5
NR2F1
TEAD1
BRCA1
HBP1
IRX5
SMAD5
SMAD6
SOX3
SOX6
GABPA
ELF1
ERF
ID2
GLI1
PLAGL1
TSC22D4
1
0.8
0.6
0.4
0.2
B)
Cell type gene expression ranked
TF targets
TF1
TF2
TF3
TF4
FWY Results
TF
p-value
1,2,3
<1e-3
2,3,4
1.5e-2
1,2
3e-2
3,4
2e-2
Network Representation
TF4
TF3
TF1
TF2
Combinatorial of TF target genes
FWY procedure
Stadistical significant target combinations
C)
ARNTL
SREBF1
ERF
HES5
RFX4
GABPA
TGIF1
OTX2
TGIF2
E2F5
FOXM1
TCF4
MEIS1
SOX2
ASCL1
E2F3
TCF12
TCF3
ATF5
SOX3
SOX9
SOX21
SOX5
PAX6
SOX6
D)
counts
Number TF in combination
E)
OTX2
PAX6
TGIF2
FOXM1
ERF
SOX6
RFX4
ATF5
MEIS1
TGIF1
HES5
SOX9
SOX3
GABPA
SOX21
SREBF1
SOX2
TCF3
TCF12
TCF4
E2F5
SOX5
ASCL1
E2F3
ARNTL
TF fraction in combinations
F)
Log(Weighted Degree)
Node Degree
ARNTL
E2F3
TCF3
ASCL1
E2F5
SOX2
TCF4
TCF12
SREBF1
GABPA
SOX5
SOX21
HES5
SOX9
TGIF1
RFX4
MEIS1
ATF5
ERF
TGIF2
FOXM1
G)
TCF3
TCF4
ASCL1
TCF12
SOX2
SOX5
SREBF1
ARNTL
E2F3
E2F5
TF
Targets
ALDH1A1
TH
BMP7
DVL2
NEUROD1
NFE2L1
FOXA1
CTNNB1
GSK3B
TCF12
NR4A2
HDAC2
IGF1R

Figure 5

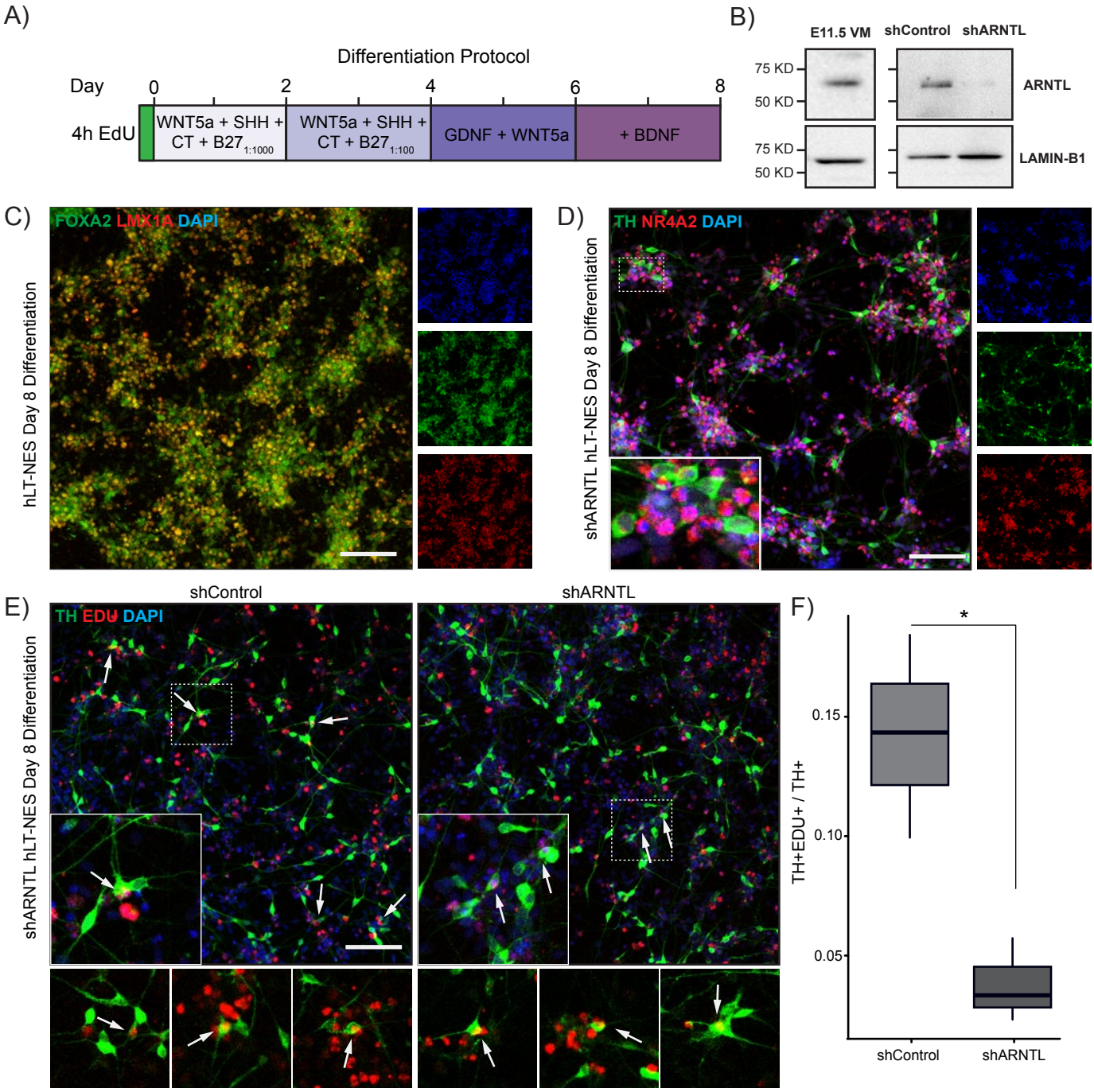

A)
Differentiation Protocol
Day
0
2
4
6
8
4h EdU
WNT5a + SHH + CT + B27 1:1000
WNT5a + SHH + CT + B27 1:100
GDNF + WNT5a
+ BDNF
B)
E11.5 VM
shControl
shARNTL
75 KD
50 KD
ARNTL
LAMIN-B1
C)
FOXA2 LMX1A DAPI
hLT-NES Day 8 Differentiation
D)
TH NR4A2 DAPI
shARNTL hLT-NES Day 8 Differentiation
E)
shControl
shARNTL
TH EDU DAPI
shARNTL hLT-NES Day 8 Differentiation
F)
*
TH+EDU+ / TH+
0.15
0.10
0.05
shControl
shARNTL

Figure 6

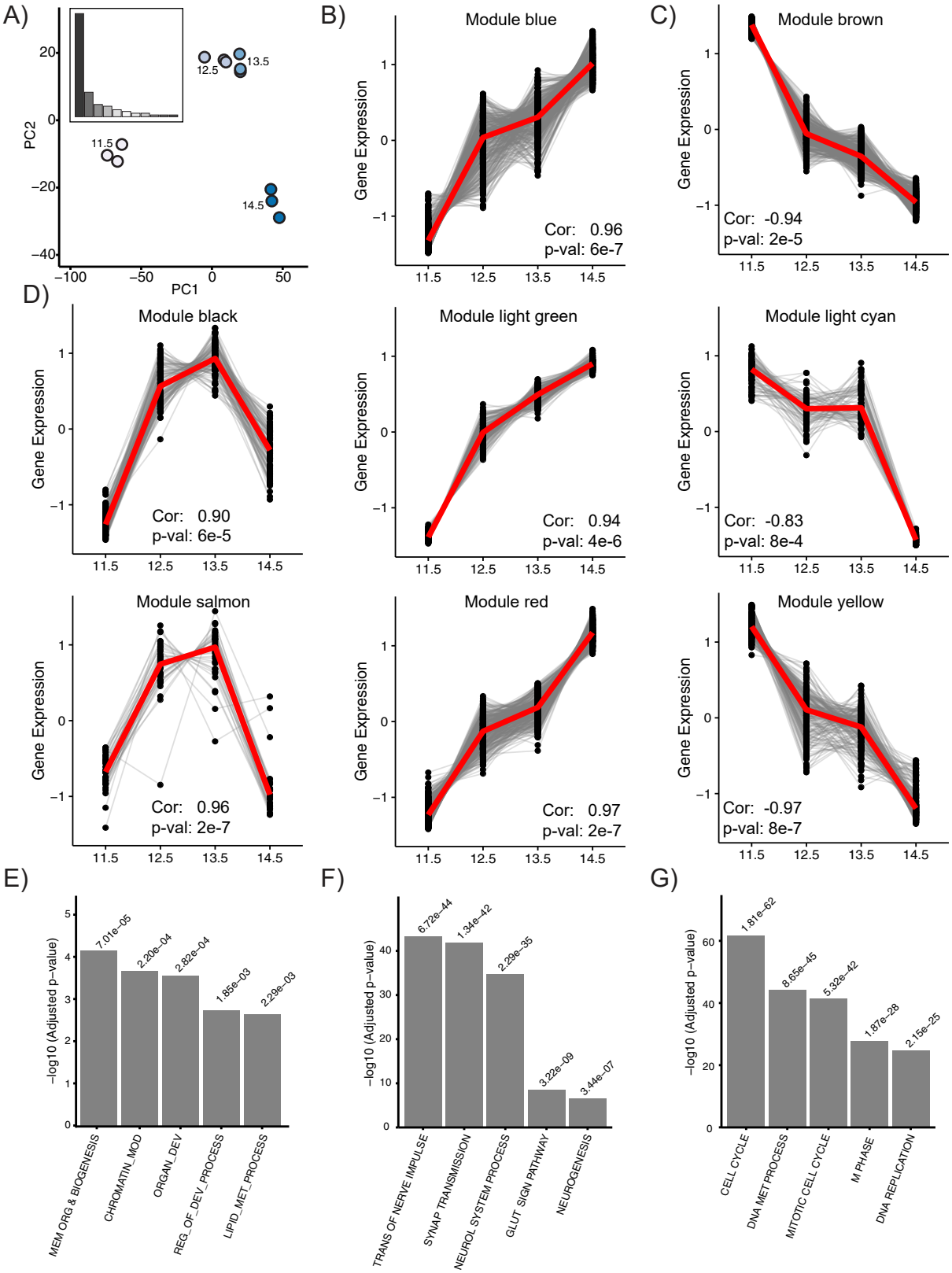

A)
PC2
PC1
12.5
13.5
11.5
14.5
B)
Module blue
Gene Expression
Cor: 0.96
p-val: 6e-7
C)
Module brown
Cor: -0.94
p-val: 2e-5
D)
Module black
Cor: 0.90
p-val: 6e-5
Module light green
Cor: 0.94
p-val: 4e-6
Module light cyan
Cor: -0.83
p-val: 8e-4
Module salmon
Cor: 0.96
p-val: 2e-7
Module red
Cor: 0.97
p-val: 2e-7
Module yellow
Cor: -0.97
p-val: 8e-7
E)
−log10 (Adjusted p−value)
7.01e−05
2.20e−04
2.82e−04
1.85e−03
2.29e−03
MEM ORG & BIOGENESIS
CHROMATIN_MOD
ORGAN_DEV
REG_OF_DEV_PROCESS
LIPID_MET_PROCESS
F)
6.72e−44
1.34e−42
2.29e−35
3.22e−09
3.44e−07
TRANS OF NERVE IMPULSE
SYNAP TRANSMISSION
NEUROL SYSTEM PROCESS
GLUT SIGN PATHWAY
NEUROGENESIS
G)
1.81e−62
8.65e−45
5.32e−42
1.87e−28
2.15e−25
CELL CYCLE
DNA MET PROCESS
MITOTIC CELL CYCLE
M PHASE
DNA REPLICATION

Figure S1

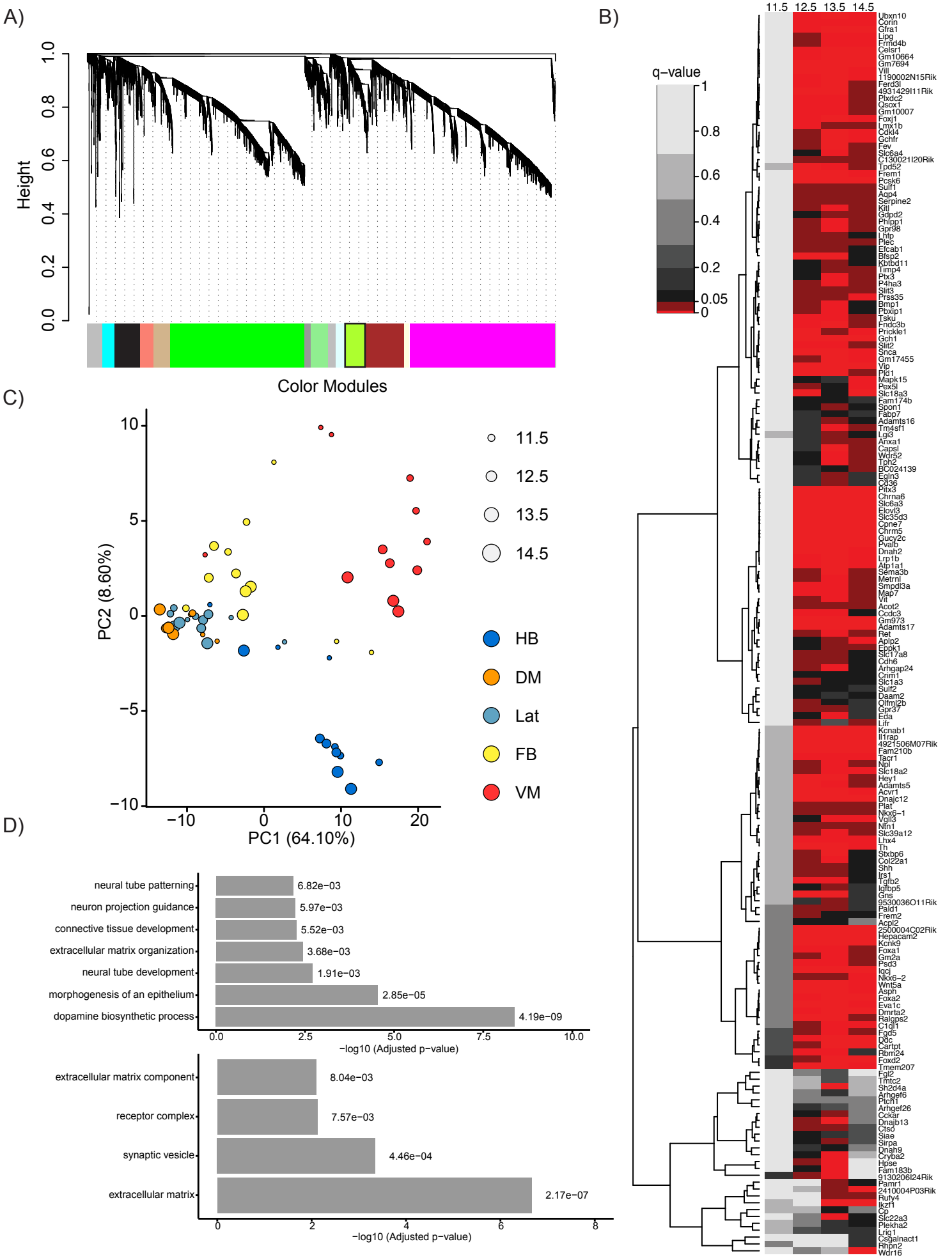

A)
Height
1.0
0.8
0.6
0.4
0.2
0.0
Color Modules
B)
11.5
12.5
13.5
14.5
q−value
1
0.8
0.6
0.4
0.2
0.05
0
C)
PC2 (8.60%)
PC1 (64.10%)
11.5
12.5
13.5
14.5
HB
DM
Lat
FB
VM
D)
neural tube patterning
neuron projection guidance
connective tissue development
extracellular matrix organization
neural tube development
morphogenesis of an epithelium
dopamine biosynthetic process
6.82e−03
5.97e−03
5.52e−03
3.68e−03
1.91e−03
2.85e−05
4.19e−09
−log10 (Adjusted p−value)
extracellular matrix component
receptor complex
synaptic vesicle
extracellular matrix
8.04e−03
7.57e−03
4.46e−04
2.17e−07
−log10 (Adjusted p−value)

Figure EV2

A)

B)

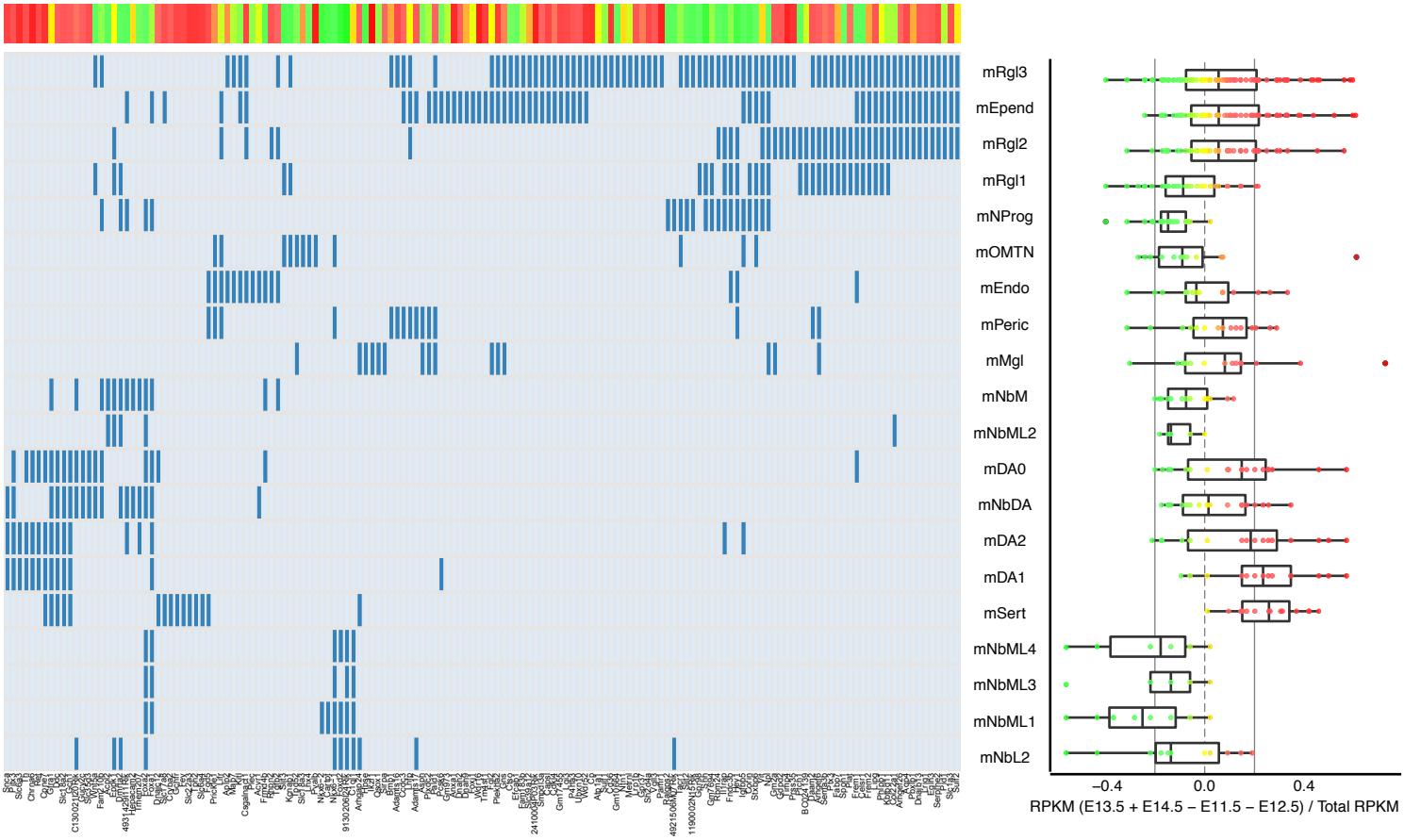

mRgl3
mEpend
mRgl2
mRgl1
mNProg
mOMTN
mEndo
mPeric
mMgl
mNbM
mNbML2
mDA0
mNbDA
mDA2
mDA1
mSert
mNbML4
mNbML3
mNbML1
mNbL2
−0.4
0.0
0.4
RPKM (E13.5 + E14.5 − E11.5 − E12.5) / Total RPKM

Figure S3

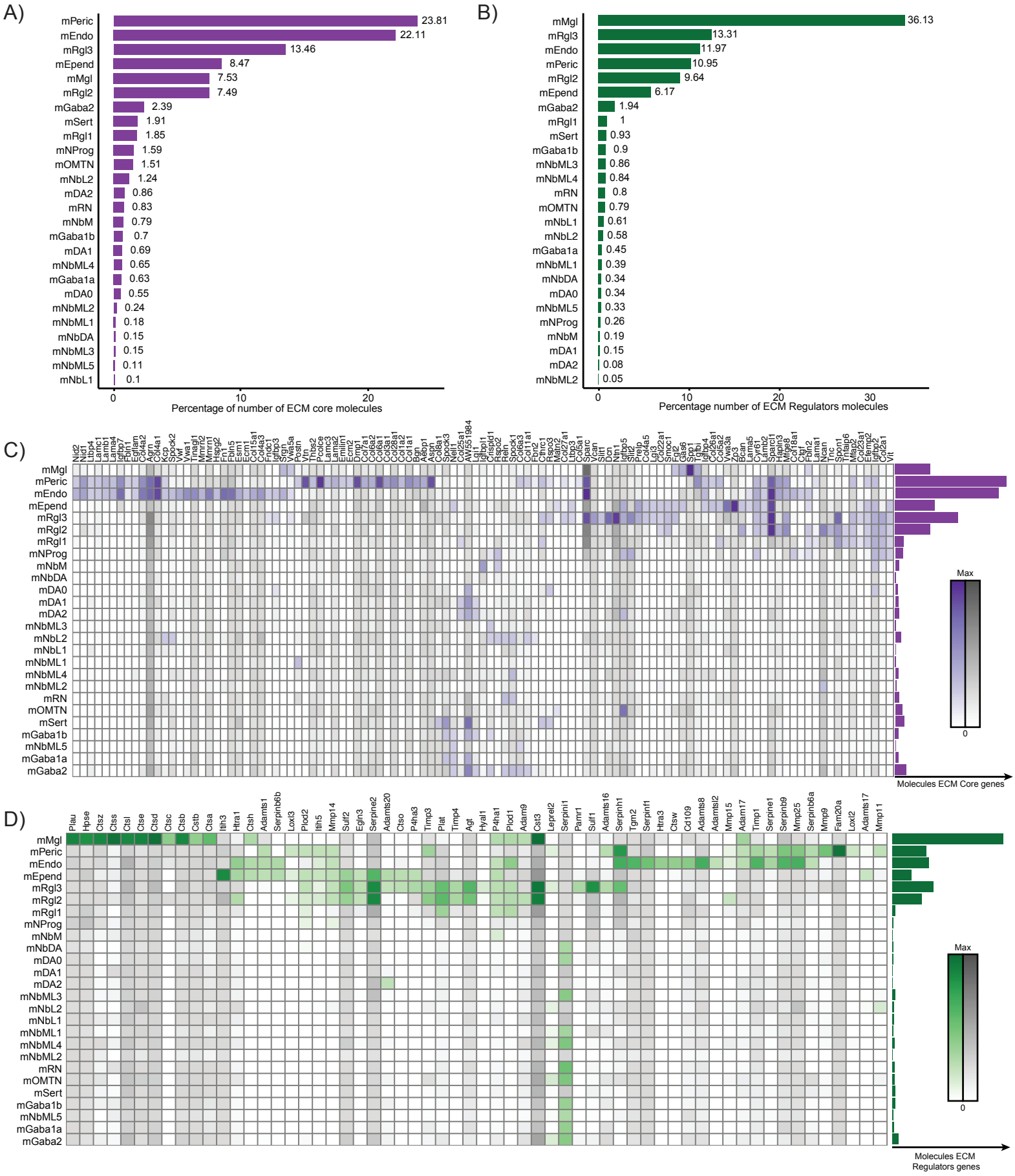

A)
mPeric 23.81
mEndo 22.11
mRgl3 13.46
mEpend 8.47
mMgl 7.53
mRgl2 7.49
mGaba2 2.39
mSert 1.91
mRgl1 1.85
mNProg 1.59
mOMTN 1.51
mNbL2 1.24
mDA2 0.86
mRN 0.83
mNbM 0.79
mGaba1b 0.7
mDA1 0.69
mNbML4 0.65
mGaba1a 0.63
mDA0 0.55
mNbML2 0.24
mNbML1 0.18
mNbDA 0.15
mNbML3 0.15
mNbML5 0.11
mNbL1 0.1
0
10
20
Percentage of number of ECM core molecules
B)
mMgl 36.13
mRgl3 13.31
mEndo 11.97
mPeric 10.95
mRgl2 9.64
mEpend 6.17
mGaba2 1.94
mRgl1 1
mSert 0.93
mGaba1b 0.9
mNbML3 0.86
mNbML4 0.84
mRN 0.8
mOMTN 0.79
mNbL1 0.61
mNbL2 0.58
mGaba1a 0.45
mNbML1 0.39
mNbDA 0.34
mDA0 0.34
mNbML5 0.33
mNProg 0.26
mNbM 0.19
mDA1 0.15
mDA2 0.08
mNbML2 0.05
0
10
20
30
Percentage number of ECM Regulators molecules
C)
Max
0
Molecules ECM Core genes
D)
Max
0
Molecules ECM Regulators genes

Figure S4

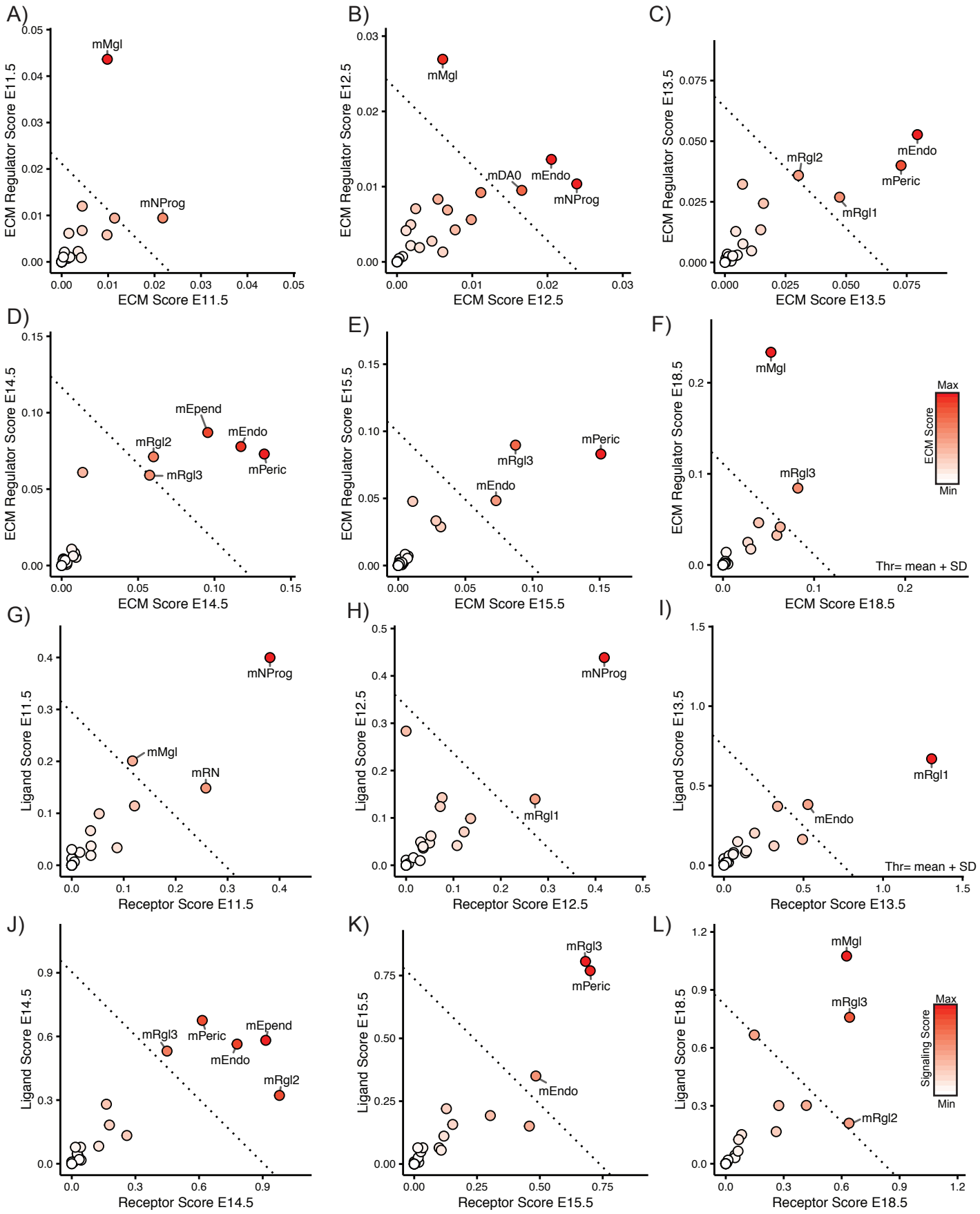

A)
ECM Regulator Score E11.5
ECM Score E11.5
mMgl
mNProg
B)
ECM Regulator Score E12.5
ECM Score E12.5
mMgl
mDA0
mEndo
mNProg
C)
ECM Regulator Score E13.5
ECM Score E13.5
mRgl2
mEndo
mPeric
mRgl1
D)
ECM Regulator Score E14.5
ECM Score E14.5
mEpend
mEndo
mRgl2
mPeric
mRgl3
E)
ECM Regulator Score E15.5
ECM Score E15.5
mPeric
mRgl3
mEndo
F)
ECM Regulator Score E18.5
ECM Score E18.5
mMgl
mRgl3
Max
ECM Score
Min
Thr= mean + SD
G)
Ligand Score E11.5
Receptor Score E11.5
mNProg
mMgl
mRN
H)
Ligand Score E12.5
Receptor Score E12.5
mNProg
mRgl1
I)
Ligand Score E13.5
Receptor Score E13.5
mRgl1
mEndo
Thr= mean + SD
J)
Ligand Score E14.5
Receptor Score E14.5
mRgl3
mPeric
mEpend
mEndo
mRgl2
K)
Ligand Score E15.5
Receptor Score E15.5
mRgl3
mPeric
mEndo
L)
Ligand Score E18.5
Receptor Score E18.5
mMgl
mRgl3
Max
Signaling Score
Min
mRgl2

Figure S5

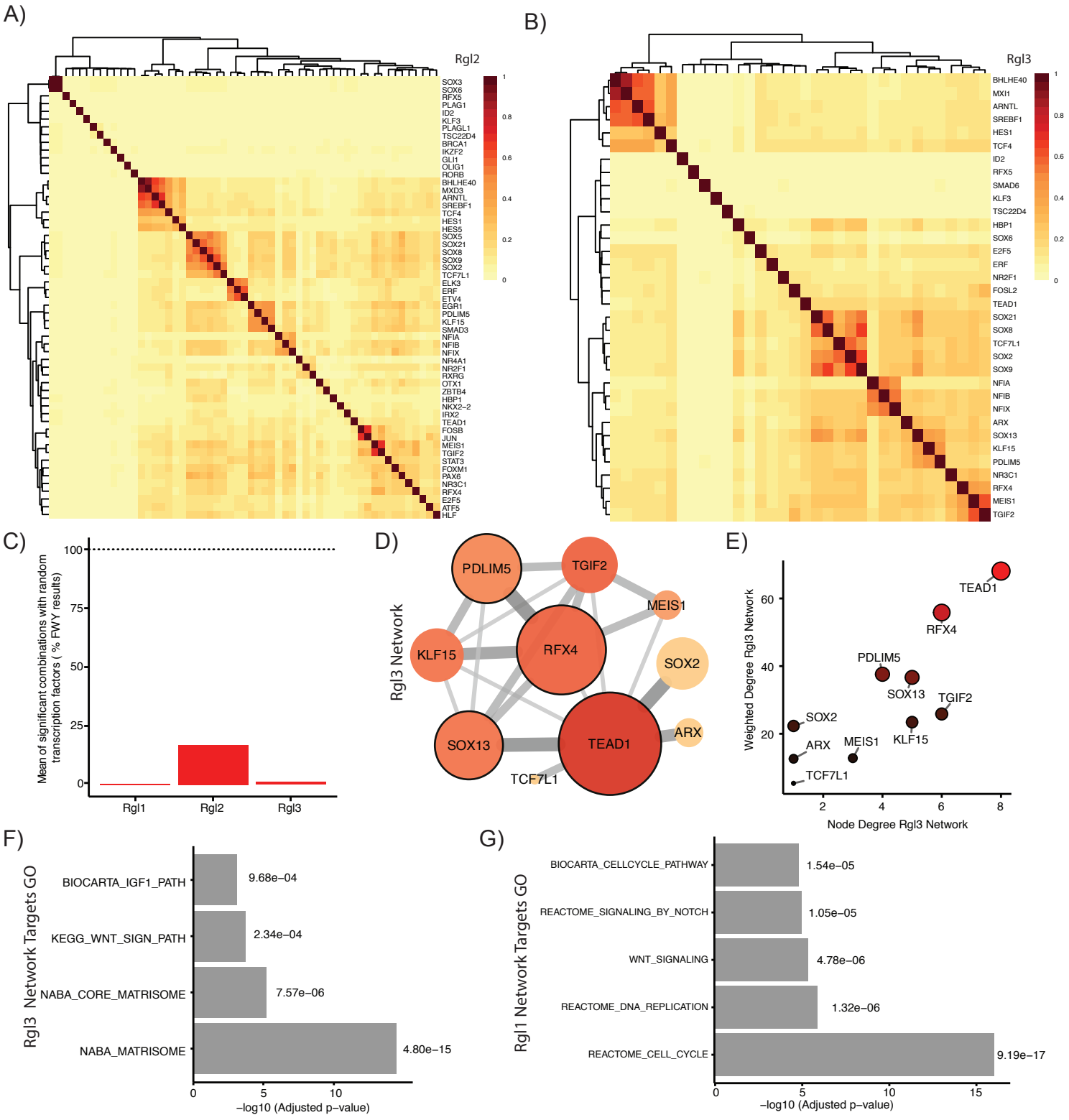

A)
Rgl2
SOX3
SOX6
RFX5
PLAG1
ID2
KLF3
PLAGL1
TSC22D4
BRCA1
IKZF2
GLI1
OLIG1
RORB
BHLHE40
MXD3
ARNTL
SREBF1
TCF4
HES1
HES5
SOX5
SOX21
SOX8
SOX9
SOX2
TCF7L1
ELK3
ERF
ETV4
EGR1
PDLIM5
KLF15
SMAD3
NFIA
NFIB
NFIX
NR4A1
NR2F1
RXRG
OTX1
ZBTB4
HBP1
NKX2-2
IRX2
TEAD1
FOSB
JUN
MEIS1
TGIF2
STAT3
FOXM1
PAX6
NR3C1
RFX4
E2F5
ATF5
HLF
1
0.8
0.6
0.4
0.2
0
B)
Rgl3
BHLHE40
MXI1
ARNTL
SREBF1
HES1
TCF4
ID2
RFX5
SMAD6
KLF3
TSC22D4
HBP1
SOX6
E2F5
ERF
NR2F1
FOSL2
TEAD1
SOX21
SOX8
TCF7L1
SOX2
SOX9
NFIA
NFIB
NFIX
ARX
SOX13
KLF15
PDLIM5
NR3C1
RFX4
MEIS1
TGIF2
1
0.8
0.6
0.4
0.2
0
C)
Mean of significant combinations with random transcription factors (% FWY results)
100
75
50
25
0
Rgl1
Rgl2
Rgl3
D)
Rgl3 Network
PDLIM5
TGIF2
MEIS1
KLF15
RFX4
SOX2
SOX13
TEAD1
ARX
TCF7L1
E)
Weighted Degree Rgl3 Network
60
40
20
TEAD1
RFX4
PDLIM5
SOX13
TGIF2
SOX2
ARX
MEIS1
KLF15
TCF7L1
2
4
6
8
Node Degree Rgl3 Network
F)
Rgl3 Network Targets GO
BIOCARTA_IGF1_PATH
9.68e−04
KEGG_WNT_SIGN_PATH
2.34e−04
NABA_CORE_MATRISOME
7.57e−06
NABA_MATRISOME
4.80e−15
0
5
10
−log10 (Adjusted p−value)
G)
Rgl1 Network Targets GO
BIOCARTA_CELLCYCLE_PATHWAY
1.54e−05
REACTOME_SIGNALING_BY_NOTCH
1.05e−05
WNT_SIGNALING
4.78e−06
REACTOME_DNA_REPLICATION
1.32e−06
REACTOME_CELL_CYCLE
9.19e−17
0
5
10
15
−log10 (Adjusted p−value)

Figure S6

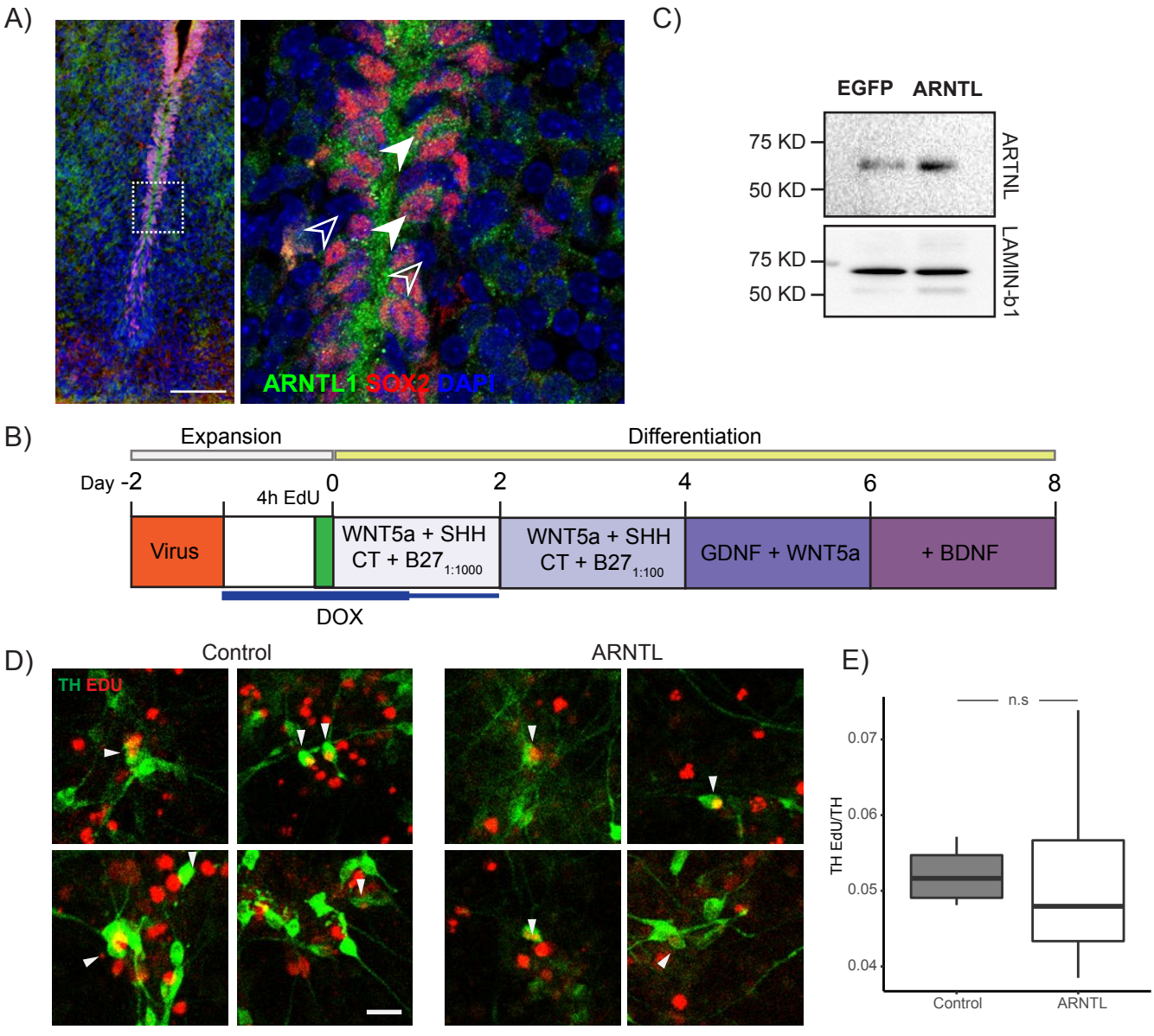

A)
ARNTL1 SOX2 DAPI
C)
EGFP ARNTL
75 KD
50 KD
ARTNL
75 KD
50 KD
LAMIN-b1
B)
Expansion
Differentiation
Day -2
4h EdU
0
2
4
6
8
Virus
WNT5a + SHH
CT + B27 1:1000
WNT5a + SHH
CT + B27 1:100
GDNF + WNT5a
+ BDNF
DOX
D)
Control
ARNTL
TH EDU
E)
n.s
TH EdU/TH
0.07
0.06
0.05
0.04
Control
ARNTL

Figure S7