Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	05/03/2021
Yajun He	Date

Computational Construction of Gene Regulatory Network and Trajectory of Developing Midbrain Dopaminergic Neurons Perturbed by Insecticides

By

Yajun He

Master of Science in Public Health

Environmental Health

Qiang Zhang, M.D, Ph.D.

Committee Chair

Computational Construction of Gene Regulatory Network and Trajectory of Developing Midbrain Dopaminergic Neurons Perturbed by Insecticides

By

Yajun He

Bachelor of Science

Central South University

2018

Thesis Committee Chair: Qiang Zhang, M.D, Ph.D.

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Environmental Health

2021

Abstract

Computational Construction of Gene Regulatory Network and Trajectory of Developing Midbrain Dopaminergic Neurons Perturbed by Insecticides

By

Yajun He

Background: Embryonic exposure to environmental chemicals, such as insecticides, can cause developmental defects that increase individuals' susceptibility to neurological diseases in both childhood and adulthood. Abnormal development of the midbrain dopamine circuit appears to underlie several neurological disorders, including autism spectrum disorder and attention deficit hyperactivity disorder. Epidemiological evidence has identified associations between exposure to insecticides during critical periods of brain development and neurological deficits. However, how exactly midbrain neurodevelopment is disrupted by insecticides is poorly understood.

Objective: As the development of dopaminergic (DA) neurons is underpinned by an elaborate network of transcription factors, we aim to characterize and construct the gene regulatory network (GRN) that orchestrates DA neural differentiation and migration to understand their developmental trajectories that may be perturbed by insecticides.

Method: Integrated bioinformatics and machine-learning approaches were used to analyze existing single-cell RNA sequencing data of developing mouse midbrain (La Manno et al., 2016). We firstly used R package Seurat to conduct quality control, dimension reduction, clustering, and differential expression analysis to obtain DA neuron subtypes. Then, we used pyScenic to generate a co-expression matrix according to Spearman correlation, and motif enrichment analysis using RcisTarget to obtain pruned regulons and TFs-genes regulation network. Lastly, we used scVelo to calculate RNA velocity and latent time to derive the developmental trajectories.

Results: 1901 cells were grouped into 15 cell types and visualized in 2D t-SNE and 3D UMAP map. Cell types were annotated according to the expression profile of marker genes. Immature DA neurons were sub-clustering into 4 types: NbDA, DA0, DA1 and DA2. pyScenic's output from 9 selected cell types gave us a complex network containing 603 transcription factors, 13691 target genes and 52538 edges. After cross referencing with literature, an organized, minimal network was obtained with 25 transcription factors. RNA velocity analysis produced trajectories of DA lineage from Prog to NbM, NbDA, DA0, DA1 and DA2.

Conclusion: We characterized the variety of developing midbrain DA neurons and derived a minimal GRN that underpins the normal midbrain DA development. Such GRN can be used to model the aberrations of DA neuron developmental trajectory by insecticides in future.

Computational Construction of Gene Regulatory Network and Trajectory of Developing Midbrain Dopaminergic Neurons Perturbed by Insecticides

By

Yajun He

Bachelor of Science

Central South University

2018

Thesis Committee Chair: Qiang Zhang, M.D, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Environmental Health

2021

Acknowledgements

I would first like to express my sincere gratitude to my thesis and faculty advisor Dr. Qiang Zhang, for not only his guidance and valuable advice on my thesis but also his inspiration of my interest in computational toxicology during courses and research. I really appreciate his patience and kindness during each meeting.

I would also like to thank my parents for their support and encouragement through my years of study. And to my friends and little kitty Debbie who give me a lot of spiritual comfort under the quarantine of COVID-19.

Table of Contents

3. Network inference using Scenic
4. RNA velocity analysis
    a. bioinformatics pipeline to obtain loom file
    b. scVelo for RNA velocity, latent time and parameter calculation
5. Integrating scenic network with literature

Results

1. Identification of neural cell types in developing mouse midbrain
2. Inference of TF network in developing mouse midbrain
3. mDA developmental trajectories

Discussion

Reference

Figures

Tables

Introduction

1. Anatomical location, projection and function of midbrain DA neuron

The majority of dopaminergic (DA) neuronal population of the mammalian brain is located in the ventral midbrain (VM) (Bjorklund et al., 2007). Dopaminergic neurons can release dopamine, a catecholaminergic neurotransmitter. These neurons are characterized by the presence of tyrosine hydroxylase (TH), which is the rate-limiting enzyme in the synthesis of catecholamines (Stern, 1983). Midbrain dopaminergic (mDA) neurons regulate several brain functions, such as motor control, reward behavior and cognition, and are associated with many brain diseases. mDA neurons are categorized into three groups located in three different nuclei: the substantia nigra pars compacta (SNc, known as the A9 group), the ventral tegmental area (VTA, A10 group) and the retrorubral field (RrF, A8 group) (Stern, 1986). The SNc neurons primarily project to the dorsal striatum via the nigrostriatal pathway, and functionally regulate voluntary movement. Degeneration of the SNc DA neurons is the main cause of the motor symptoms observed in Parkinson's disease (PD) (Horowski et al., 1995). VTA and RrF neurons project to the limbic system and prefrontal cortex to form the mesolimbic and mesocortical pathway. These neurons regulate reward behavior and cognitive function. Impairment of DA projection to the prefrontal cortex is related to schizophrenia and attention deficit hyperactivity disorder (ADHD) (Grace, 2016; Hadley et al., 2014).

2. Midbrain DA neurodevelopment process

2.1 Morphological development

The development of mDA neurons in mice starts around embryonic day 8 (E8) in the ventral midbrain followed by lineage specification, localization, phenotypic characterization, and synaptic

connections (Brignani & Pasterkamp, 2017; G. La Manno et al., 2016; Smidt, 2017; Smits, Burbach, & Smidt, 2006). On day E8, radial glial cells in the midbrain floor plate of the neural tube give rise to DA neural progenitor cells. Proliferating DA progenitor cells then differentiate into postmitotic DA neuroblasts. On E10.5 the DA neuroblasts migrate radially from the ventricular zone to the marginal zone, where they mature into DA neurons. Mature DA neurons then migrate laterally to their final destinations, where they populate into three distinct nuclei: SNc, VTA, and RrF. Thereafter these neurons undergo axonal growth and project rostrally to other areas of the developing brain. By E13 and E15, the first dopamine axons appear in the striatum and prefrontal cortex, respectively.

2.2 Molecular network underlying the development

Each of the major developmental processes is orchestrated by a number of morphogens and mitogens secreted by the surrounding tissues and the developing midbrain, such as sonic hedgehog (SHH), fibroblast growth factor 8 (FGF8) and wingless-related MMTV integration site 1 (WNT1) glycoprotein. SHH, FGF8 and WNT1 are the key secretory factors in the induction of midbrain floor plate and basal plate where the respective knockout mice have a severely reduced midbrain (Chi, Martinez, Wurst, & Martin, 2003; Chiang et al., 1996; Meyers, Lewandoski, & Martin, 1998). In response to these morphogenesis clues, a set of transcription factors (TF) play a vital role in determining the cellular fate, driving DA lineage specification, proliferation, differentiation, and migration. These TFs include Foxa2, Otx2, Lmx1, Sox6, Nr4a2, Pitx3, En1, Pbx1, etc., which are dynamically expressed at various stages of mDA development (Brignani & Pasterkamp, 2017; G. La Manno et al., 2016). Some TFs, such as Foxa2, Otx2, and Lmx1a/b, are expressed in neural progenitor cells to first establish the mDA lineage, which then either persist or turn off in further differentiated cells (Deng et al., 2011; Joyner, Liu, & Millet, 2000; Kittappa, Chang, Awatramani,

& McKay, 2007; Stott et al., 2013; Vernay et al., 2005). Some TFs, such as Nr4a2 and Neurog2 appear first in neuroblasts (Kele et al., 2006; Zetterstrom et al., 1997). Others, such as Pitx3 do not appear until in the late maturing mDA neurons (Nunes, Tovmasian, Silva, Burke, & Goff, 2003). In mature mDA neurons, Nr4a2 and Pitx3 drive the transcription of functional genes that encode DA-specific enzymes (TH, AADC), transporters (DAT, VMAT2), and receptors (DR2) required for DA neurotransmission (Martinat et al., 2006).

3. Effects of insecticides on midbrain DA neurodevelopment

3.1 Toxicity phenotype: sensitive time window

Many DA-related neurological diseases have developmental origin, in which the neurogenesis of mDA neurons is disrupted in the prenatal or postnatal stages. Among the risk factors, in utero exposure to environmental insecticides is of major concern (Grandjean & Landrigan, 2014; J. R. Roberts & Karr, 2012). Insecticides consist of a wide range of chemicals that are developed to purposely disrupt neuronal targets in insects (Casida & Bryant, 2017). Three major chemical classes make up 85% of the insecticides used in the United States: organochlorines, organophosphates, and pyrethroids. Routinely applied in agricultural and residential settings, they can accumulate in human bodies and transfer to the developing fetus via the placenta. A wealth of epidemiological studies have indicated that pre- and postnatal exposure to these insecticides is associated with increased risk of mDA-related neurobehavioral disorders in childhood, such as autism spectrum disorders (ASD) and attention deficit hyperactivity disorder (ADHD) (Wagner-Schuman et al., 2015). Animal studies have suggested that exposure to insecticides during the critical periods of mDA neurodevelopment can result in similar neurological disorders (Jason R Richardson et al., 2015).

3.2 Molecular changes

Over the last several years, studies have been extensively focused on determining the neurological targets of specific insecticides and their contribution to neurological disease (Caudle, Guillot, Lazo, & Miller, 2012; Caudle, Richardson, Wang, & Miller, 2005; Hatcher et al., 2007). Developmental exposure to insecticides, including dieldrin, heptachlor, endosulfan, and deltamethrin severely alter the development and function of the dopamine circuit in male and female offspring. These alterations also appear to underlie observed behavioral impairments, including deficits in social interaction, locomotor activity, and learning and memory (Caudle et al., 2005; Elwan, Richardson, Guillot, Caudle, & Miller, 2006; Richardson et al., 2006; Richardson et al., 2008; Wilson, Shapiro, Bradner, & Caudle, 2014). These findings are critical as alterations in the dopamine circuit have been identified in a variety of neurodevelopmental diseases (Di Martino et al., 2011; Greene et al., 2011; Kohls et al., 2013; Kohls, Yerys, & Schultz, 2014; Langen et al., 2009; Nakamura et al., 2010; Scott-Van Zeeland et al., 2010).

Organochlorine: Organochlorine insecticides, such as dieldrin, heptachlor, and endosulfan, affect sodium channels, as well as calcium signaling in the neuron. Although these insecticides target ubiquitous neuronal elements, each have been shown to significantly damage the dopamine system. Exposure to dieldrin or heptachlor throughout gestation and lactation elicited elevations in the expression of striatal DAT and VMAT2 in both male and female offspring at 3 months of age (Caudle et al., 2005; Richardson et al., 2006; Richardson et al., 2008). These elevations appear to be mediated by the transcription factors, Nurr1 and Pitx3, which are highly involved in conferring the dopaminergic phenotype to mature dopamine neurons. IMale offspring developmentally exposed to endosulfan demonstrated a significant reduction in striatal DAT and TH expression (Wilson et al., 2014). Damage to these circuits resulted in significant impairments

in social interactions in these mice. These findings are in line with the epidemiological data showing developmental exposure to endosulfan to be associated with an elevated incidence of ASD in children (E. M. Roberts et al., 2007).

Pyrethroids: Pyrethroid insecticides are found in two different flavors: Type I pyrethroids, which include permethrin, resmethrin, and allethrin, primarily exert their neurotoxic effects by disrupting the function of sodium channels, leading to neuronal hyperexcitation. Type II pyrethroids, which include deltamethrin, cypermethrin, and cyhalothrin, also alter the function of sodium channels, but simultaneously serve as inhibitors of the GABA(A) receptor, resulting in elevated neuronal excitation. Exposure to pyrethroid insecticides appears to have similar effects on the striatal dopamine system as those seen with organochlorines. Developmental exposure to deltamethrin causes significant alterations to dopamine handling and transport in the striatum of offsprings exposed during gestation and lactation (J. R. Richardson et al., 2015). Specifically, while both male and female offsprings demonstrated a dose-dependent elevation in striatal DAT, only the male mice exhibited an alteration in concentrations of extracellular dopamine and behavioral impairments associated with the striatum and ADHD (J. R. Richardson et al., 2015).

Organophosphates: Typical organophosphate compounds are chlorpyrifos and malathion. The specificity of organophosphates to the acetylcholine system through the targeted inhibition of the acetylcholine degradation enzyme, acetylcholinesterase, would suggest an exclusive effect on the cholinergic pathway. However, the dopamine circuit has also been shown to be a target following developmental exposure (Torres-Altoro et al., 2011). While these findings were generated following exposure to adult animals, exposure of animals during critical periods of neurodevelopment have also been shown to cause reductions in dopamine signaling (Slotkin et al., 2009). Additionally, many of the effects on the dopamine system were observed below the

threshold concentration or dose for inhibition of acetylcholinesterase, suggesting that these findings were independent of the specific cholinergic target of organophosphate insecticides.

Taken together, these findings highlight the dopamine circuit as a critical target of exposure to organochlorine, organophosphate, and pyrethroid insecticides. Moreover, alterations to this circuit, especially during critical periods of neurodevelopment, delineate an important neuronal pathway that may underlie the epidemiological findings of exposure to these compounds and neurodevelopmental deficits. While these findings provided undisputed evidence for involvement of molecular disruptions, how exactly the developing trajectories of mDA neurons are altered by insecticides and how the composition of the final populations of the mDA neurons in SNc, VTA, and RrF is affected are still largely unknown.

4. Understanding the mechanism of normal mDA neuronal development that can be perturbed by insecticides

4.1 Molecular network and developmental trajectory

While many of the molecular details of the TF network in the developing midbrain are worked out over the decades, how these TFs work together as a whole that underpins the complex mDA developmental process is poorly understood. Neural development involves neural stem cells bifurcating into and transitioning through multiple intermediate and terminal cell fates in a physiologically irreversible fashion. A prevailing premise, based on nonlinear dynamical systems theory, purports that stable cell types or subtypes correspond to stable steady states (attractors) in a high-dimensional molecular state landscape underpinned by the TF network (Enver, Pera, Peterson, & Andrews, 2009; Huang, Eichler, Bar-Yam, & Ingber, 2005; Mojtahedi et al., 2016). Albeit with caveats, the Waddington epigenetic landscape is a useful visual aid for such attractor

landscape during development (Waddington, 1957). Cell state transitions through a developmental trajectory can be directed by transient warping of the landscape by local morphogens. While morphogens provide instructive signals, the cellular fates that ensue are also modulated by stochastic factors resulting from gene expression noise (Kaern, Elston, Blake, & Collins, 2005; Raj & van Oudenaarden, 2008). The stochastic fluctuation of mRNA and protein levels has been well documented for its crucial role in determining the timing and probability of the fate of individual cells, and consequently, the proportion of cell types and subtypes at a given time point during development (Balazsi, van Oudenaarden, & Collins, 2011; Symmons & Raj, 2016; Zhang, Bhattacharya, Kline, et al., 2010). A number of studies have modelled and experimentally investigated developmental TF networks as multi-stable dynamical systems, including those operating in embryonic stem cells, hematopoietic stem cells, and B cells (Bhattacharya et al., 2010; Chickarmane, Troein, Nuber, Sauro, & Peterson, 2006; Duff, Smith-Miles, Lopes, & Tian, 2012; Harris, Pargett, Sutcliffe, Umulis, & Ashe, 2011; Zhang, Bhattacharya, & Andersen, 2010; Zhang et al., 2013).

4.2 How insecticides may perturb the network and trajectory

So it is highly likely that the mDA developmental process, starting from radial glial cells, traversing through DA neural progenitor cells, DA neuroblasts, and finally to mature, terminally differentiated DA neurons and their subtypes, is an orchestrated event occurring along the mDA attractor landscape that is underpinned by the specific TF network (Blaess & Ang, 2015; S. V. Hegarty, A. M. Sullivan, & G. W. O'Keeffe, 2013). Environmental toxicants like insecticides may directly or indirectly perturb this TF network, resulting in transiently or permanently altered landscapes. The landscape may change in such a way that the movement of DA lineage cells is either accelerated or decelerated at some steps of the developmental trajectories, resulting in an

aberrant number of mature DA neurons or disproportional populations of DA neuron subtypes. It is also possible that the alteration of the landscape results in new, aberrant trajectories that lead cells to alternative attractor states that are never or rarely visited in normal developmental conditions. Therefore, we hypothesize that developmental exposure to insecticides can alter the transcriptional network that underpins midbrain dopamine neurodevelopment such that (1) the developing neurons follow an abnormal developmental trajectory leading to aberrant DA neuron subtypes, and/or (2) the probability of developing DA neurons moving along the normal trajectories is altered resulting in aberrant proportions of DA neurons in each subtype.

4.3 A computational modeling approach is necessary to map and simulate the TF network and trajectory

To test the hypothesis efficiently, it requires integrated multidisciplinary systems biology approaches – a combination of data-driven, top-down approach examining the TF network globally and hypothesis-driven, bottom-up approach constructing dynamical models of the network piece by piece. Assessing the molecular profile or state of the TF network in individual cells demands a data-intensive, high-coverage method where all the TFs in the network, known and unknown, can be measured at once at single-cell levels.

4.4 Utilizing scRNA-seq to map the TF network and facilitate modeling

In the past few years, the emerging single-cell RNA sequencing (scRNA-seq) technology has made it possible to distinguish and classify cells based on their global transcriptomic signatures including neural cells (G. La Manno et al., 2016; Sagner et al., 2018). More recently, the NIH's Brain Research through Advancing Innovative Neurotechnologies (BRAIN) has started the BRAIN Initiative Cell Census Network (BICCN) aiming to map all neural cells in the brain (Ecker et al.,

2017). ScRNA-seq studies of the developing midbrain can produce data from which several pieces of information can be extracted and fed into the bottom-up dynamical modeling approach.

- **scRNA-seq could inform neuron subtypes during DA development**

ScRNA-seq studies will allow high-resolution profiling of the mDA neurons, thus facilitating the counting of different cell subtypes and evaluating their proportions that can be potentially altered by insecticide exposure. A couple of recent single-cell studies have identified several mDA neuron subtypes in the developing and postnatal midbrain tissues in mice, which were concealed in traditional methods based on neuron location, morphology, function, and specific markers (Hook et al., 2018; G. La Manno et al., 2016; Poulin et al., 2014). The scRNA-seq approach will also allow us to discover whether new DA neuron subtypes can emerge from insecticide exposure. The number of identified mDA neuron subtypes will provide information on the number of possible attractor states of the TF network model.

- **scRNA-seq could infer network structure**

scRNA-seq technology has made it possible to infer the gene regulatory networks (GRN) that control cellular differentiation and drive the transitions from one cell type to another. Bioinformatics methods such as SCENIC and GRISLI can be applied to identify the coregulated modules of genes and GRN using gene-to-gene expression correlations (Aibar et al., 2017; Aubin-Frankowski & Vert, 2020).

- **scRNA-seq can infer pseudo-time and latent time**

Novel, emerging machine-learning tools such as Monocle and RNA velocity have enabled the identification of developmental pseudotime and latent time of each cell assayed with scRNA-seq

and are able to order the heterogeneous cells in tissue samples along a developmental trajectory (Hook et al., 2018; G. La Manno et al., 2016). This derived information will not only provide visualization of the potentially altered trajectory induced by insecticide exposure but also provide secondary data to guide dynamical modeling of the TF network. RNA velocity could also infer parameters such as transcription rate and degradation rate constant that can be used in dynamic modeling of the TF network.

Method

1. Mouse developing midbrain scRNA-seq Data source

E11.5 - E18.5 mouse midbrain scRNA-seq data was downloaded from Gene Expression Omnibus (GEO) under the series number GSE76381 from a recently published paper (G. La Manno et al., 2016). A total of 2486 raw sequencing FASTQ files and a molecule count matrix with 1901 cells X 24378 genes were obtained.

2. Analysis using Seurat

We used Seurat (v3.2.2) in R (v3.6.2) to distinguish different cell types (Butler, Hoffman, Smibert, Papalexi, & Satija, 2018). Since the expression matrix we downloaded from GEO was already quality controlled by the original authors, the data was first normalized using the LogNormalize method. Feature counts for each cell are divided by the total counts for that cell and multiplied by the scale.factor set as 10000, then a natural-log was performed. After calculating the cell-to-cell variation of genes, a subset of 3000 genes that exhibit high variation in the dataset were used in downstream analysis. After scaling, PCA was performed for linear dimension reduction. Cell clustering was performed with the "resolution" parameter set to 1.7. Non-linear dimensional reduction tools tSNE (Kobak & Berens, 2019) and UMAP (Becht et al., 2019) were then used to visualize the dataset. After gene expression analysis, we annotated cell types to each cluster according to the expression profile of marker genes listed in Table 1.

3. Network inference using Scenic

We used pySCENIC (v0.10.4) through the command line interface to derive the gene regulatory network (Aibar et al., 2017). Firstly, a co-expression matrix or regulons were generated after the

calculation of gene-to-gene Spearman correlation using GENIE3. Then, we performed motif enrichment analysis by searching the promoter of 10k bp upstream and downstream of the transcription start site (TSS) to obtain pruned regulons and the TFs-genes networks using RcisTarget. TFs-genes network was visualized in Cytoscape.

4. RNA velocity analysis

a. bioinformatics pipeline to obtain loom file

2486 original sequencing FASTQ files were converted into one loom file before the RNA velocity analysis. We firstly performed quality control on 2486 FASTQ files as follows using UMI-tools and Cutadapt: (1) removed reads whose UMI bases had a Phred score < 17; (2) removed Gs from 5', low quality 3' end bases and poly A tail; (3) removed reads whose bases are less than 25. After quality control, reads were aligned to mm10 genome using Bowtie 2 aligner. Then we converted 2486 SAM files to 2486 position-sorted BAM files and generated a loom file containing unspliced RNA by using reference genome mm10 through VeloCyto (Gioele La Manno et al., 2018).

b. scVelo for RNA velocity, latent time and parameter calculation

We used scVelo (v0.2.4) through python (v3.8) to calculate RNA velocity and latent time to derive the developmental trajectories (Bergen, Lange, Peidli, Wolf, & Theis, 2020). After obtaining the loom file which stores both spliced and unspliced counts. We firstly filtered the data by only retaining mouse embryo cells with 2000 - 26000 molecules per cell and eliminating genes that were detected at less than 4 molecules in the whole datasets. After normalization, RNA velocity and latent-time were estimated using a dynamical model and projected to a 2D UMAP map. Then the rate constants of RNA transcription, splicing and degradation were calculated.

5. Integrating scenic network with literature

In order to validate and complement SCENIC-derived TF network, we cross referenced with the reported connections reported in the literature (Arenas, Denham, & Villaescusa, 2015; Blaess & Ang, 2015; Shane V Hegarty, Aideen M Sullivan, & Gerard W O'keeffe, 2013), to obtain an organized and minimally sufficient network that can be computed in future.

Result

1. Identification of neural cell types in developing mouse midbrain

Through dimension reduction and gene expression clustering, we grouped 1901 cells into 15 clusters: GABAergic neurons (GABA1-2); mediolateral neuroblasts (NbML1-5); red nucleus (RN); lateral neuroblasts (NbL1-2); neuronal progenitor (NProg); neuroblast dopaminergic (NbDA); dopaminergic neurons (DA0-2); ependymal (Epend); radial glia-like cells (Rgl1-3); endothelial cells (Endo); medial neuroblast (NbM); pericytes (Peric); oculomotor and trochlear nucleus (OMTN); serotonergic (Sert); microglia (Mgl). These clusters of cells are largely concordant with those reported by La Manno 2016. We visualized them in 2D t-SNE (Figure 1A) and 3D UMAP (Figure 2) maps. Immature DA neurons were sub-clustered into 4 types: NbDA, DA0, DA1 and DA2 (Figure 1B). Cell types were annotated according to the expression profile of marker genes.

2. Inference of TF network in developing mouse midbrain

Since only cell types Rgl1, Rgl2, Rgl3, NProg, NbM, mNbDA, DA0, DA1, and DA2 are immediately relevant to the mDA lineage development, we only applied the SCENIC algorithm to these cells. pyScenic's output from these nine selected cell types produced a complex network containing 603 transcription factors, 13691 target genes and 52538 edges. After cross referencing with literature, an organized, minimal network was obtained with 25 transcription factors: Otx2, Foxa2, Lmx1a, Lmx1b, Gbx2, Wnt1, Ctnnb, Shh, Gli2, Gli1, Neurog2, Msx1, Pbx1, Onecut2, Nr4a2, Pitx3, En1, En2, Wnt5a, Nkx2-2, Nkx6-1, Hes1, Ascl1, Pou2f2, Gata4. Network was visualized in Cytoscape (Figure 3). From upstream to downstream, the network follows the direction of DA neuron differentiation and maturation.

3. mDA developmental trajectories

The relative abundance of unspliced over spliced RNA in the dataset is presented in Figure 4, the ratio of unspliced over spliced RNA for each gene determines its RNA velocity. Projection of velocity onto a 2D UMAP map inferred the trajectories and latent time of midbrain neural development (Figure 5), which describes the direction in which each cell moves in developmental time. Neural cells appear to diverge into two cellular outcomes: one is GABAergic neurons and the other is dopaminergic neurons. DA lineage-relevant cell types, including mNProg, mNbM, mNbDA, mDA0, mDA1 and mDA2, were further analyzed which produced a linear trajectory projection, in which mDA0 diverged from mDA1/2 (Figure 6). The expression of major marker genes for DA neurons over the predicted latent time is shown in Figure 7, including Otx2, Foxa2, Lmx1a, Neurog2, Nr4a2, Pitx3, En1, Th, and Slc6a3.

Discussion

Midbrain DA neural development follows a specific sequence of events involving dynamic expression of multiple TF genes and execution genes. Some of these genes and interactions are obligatory because their deletion of disruption will lead to developmental defects. Some of them are redundant to increase biological robustness. For the purpose of understanding and predicting the effects of herbicides on DA development, it is likely that all these relevant genes and interactions need to be considered. The existing literature only covers a part of the TF network, which needs to be complemented by information from new dataset. scRNA-seq data contains a multitude of information that can be extracted, some which can inform the network structure of the TF networks.

The minimal TF network we obtained using the SCENIC algorithm is shown in Figure 3, which captures the major TFs and regulations involved in midbrain DA development. The complex TF network comprises a variety of interconnected network motifs, including positive feedback, double-negative feedback, negative feedback, and feedforward loops. The regulation of VM patterning involves key TFs Otx2, Gbx2, Lmx1a, Lmx1b and Wnt1/b-catenin. Otx2 and Gbx2 mutually inhibit each other and form a double negative feedback loop, which determines the boundary between the midbrain and hindbrain cell fates. Feedforward loop of Otx2, Lmx1a, Lmx1b and Wnt1/b-catenin mediate the specification of midbrain floor plate. The positive feedback loop between Foxa2, Gli2 and SHH establishes the initial mDA lineage identity (Mavromatakis et al., 2011). First, SHH coming from the notochord underneath the neural tube activates Gli2 to induce Foxa2 in the neuroepithelial or radial glial cells in the floor plate (Chamberlain, Jeong, Guo, Allen, & McMahon, 2008); Foxa2 in turn induces SHH itself in these cells, which is secreted and acts in an autocrine and paracrine manner to further induce Foxa2.

This positive feedback loop leads to a potential self-sustained, stable steady state, Foxa2^{high} and SHH^{high}, to establish the mDA lineage (Kittappa et al., 2007; Stott et al., 2013). Msx1 and Neurog2 are downstream targets of Lmx1a/b which play an essential role in DA neurogenesis. They are mainly expressed in NbM from their expression profile over the predicted latent time shown in Figure 7. After neurogenesis, postmitotic DA neuroblasts migrate radially from ventricular zone to the marginal zone, and there they mature into DA neurons. Nr4a2, Pitx3 and En1/2 are crucial regulators of the differentiation of mDA neurons. In mature mDA, Nr4a2 and Pitx3 form a predominant positive feedback loop maintaining the mature DA identity and driving a suite of DA-specific enzymes (Martinat et al., 2006; Yi et al., 2014; Zetterstrom et al., 1997).

SCENIC identified many interactions which have not been reported in the literature. For example, SCENIC's output shows over 600 interactions between Pbx1 and other upstream TFs which are rarely reported in literature. Villaescusa's study shows Pbx1's role in inhibiting lateral fates during embryogenesis by repressing Onecut2 and promoting mDA development by activating Pitx3 which is consistent with our results (Villaescusa et al., 2016). In addition to Pbx1, many other interactions remain to be validated.

The initial network inferred from SCENIC is large: containing 603 transcription factors, 13691 target genes and 52538 edges. Unlike many other network inference algorithms which relies purely on gene-gene expression correlation, the SCENIC algorithm utilizes the known TF binding sites and search the enrichment of the binding motifs in the upstream promoter of the initially identified "target" genes in the regulon based on correlation score by using the RcisTraget database. This extra step screens out those correlated genes that may not be a physical target of the TFs (false positive), thus narrowing down the set of genes that are the true target of the TF. However, because scRNA-seq data are highly variable, many lowly-expressed genes may only

have a couple of copies or even lower mRNA molecules of a particular gene in a small subset of cells specifically expressing the gene, so the initial expression correlation based identification of target genes may not have those genes and interactions included (false negative). A typical example in the current study is Nr4a2, which has no associated regulon identified at all although it is a well-known TF that is responsible for the terminal phenotype of DA neurons.

Alternative GRN inference algorithms include those based on dynamical systems that are described by ordinary differential equations (ODE) such as SCODE and GRISLI ((Aubin-Frankowski & Vert, 2020; Matsumoto et al., 2017). However, these complementary approaches are not necessarily immune to the false negative problem. This is because the inferred Jacobian matrix still replies on the data of gene expression. However, the newly developed GRISLI utilizes the RNA velocity concept and is expected to perform better than those algorithms previously developed based on bulk microarray assays and transcriptomic assays. It is worth comparing SCENIC derived network with that predicted by GRISLI in future for DA neural development.

Using the RNA velocity concept and scVelo we have derived the trajectories of the DA neuron development in the mouse midbrain (Figures 5 and 6). Conceptually, the RNA velocity in the reduced 2D space is analogous to the vector field in dynamical systems, which predict the rate and direction at which the system is moving toward in the next moment. The trajectory depicted here is surprisingly consistent with the known sequence of DA neuronal development, therefore providing confidence to its validity. While the samples of the scRNA-seq data were from several developmental stages of the mouse midbrain, the latent time predicted provides information on continuous changes of gene expression during development time (Figure 7). The latent time information can be used to validate if the TF network inferred makes any sense and point out the missing links in the network that can be explored in future studies.

One of the ultimate purposes of the above bioinformatic results is to guide the development of a network model to recapitulate the transcriptional events during midbrain DA neural development and predict the consequence of insecticide perturbation. An immediate task will be to convert the network structure into a Boolean network model to assess the possible attractor state of the network that represents the different DA neuron subtypes such as NProg, NbM, NbDA, DA0, DA1 and DA2. Subsequently the network should be coded up to generate an ODE-based dynamical model that can be simulated both deterministically and stochastically. The stochastic simulation is particularly important as it can recapitulate the decision-making process during development that is often random. Lastly, such dynamic models can be incorporated into the spatial modeling platform such as CompuCell3D to simulate the DA neuronal migration during development in the midbrain.

While much effort was spent to characterize the TF network and trajectory of midbrain mDA neural development in normal conditions, we didn't get a chance to examine the situations when the network is perturbed by chemicals including insecticides. A search in the Comparative Toxicogenomic Database returned a couple of early microarray studies examining the organophosphates on mouse midbrain development (Cole et al., 2011; Moreira et al., 2010). Future work will compare the gene expression changes with the scRNA-seq data and predict the genes that are perturbed by the insecticides.

Reference

Aibar, S., González-Blas, C. B., Moerman, T., Imrichova, H., Hulselmans, G., Rambow, F., . . . van den Oord, J. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods, 14*(11), 1083-1086.

Arenas, E., Denham, M., & Villaescusa, J. C. (2015). How to make a midbrain dopaminergic neuron. *Development, 142*(11), 1918-1936.

Aubin-Frankowski, P.-C., & Vert, J.-P. (2020). Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. *Bioinformatics, 36*(18), 4774-4780.

Balazsi, G., van Oudenaarden, A., & Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell, 144*(6), 910-925. doi:10.1016/j.cell.2011.01.030

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W., Ng, L. G., . . . Newell, E. W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology, 37*(1), 38-44.

Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology, 38*(12), 1408-1414.

Bhattacharya, S., Conolly, R. B., Kaminski, N. E., Thomas, R. S., Andersen, M. E., & Zhang, Q. (2010). A bistable switch underlying B-cell differentiation and its disruption by the environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Sci, 115*(1), 51-65. doi:10.1093/toxsci/kfq035

Blaess, S., & Ang, S. L. (2015). Genetic control of midbrain dopaminergic neuron development. *Wiley Interdisciplinary Reviews: Developmental Biology, 4*(2), 113-134.

Brignani, S., & Pasterkamp, R. J. (2017). Neuronal Subset-Specific Migration and Axonal Wiring Mechanisms in the Developing Midbrain Dopamine System. *Front Neuroanat, 11*, 55. doi:10.3389/fnana.2017.00055

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol, 36*(5), 411-420. doi:10.1038/nbt.4096

Casida, J. E., & Bryant, R. J. (2017). The ABCs of pesticide toxicology: amounts, biology, and chemistry. *Toxicol Res (Camb), 6*(6), 755-763. doi:10.1039/c7tx00198c

Caudle, W. M., Guillot, T. S., Lazo, C., & Miller, G. W. (2012). Parkinson's disease and the environment: beyond pesticides. *Neurotoxicology, 33*(3), 585. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22774228

Caudle, W. M., Richardson, J. R., Wang, M., & Miller, G. W. (2005). Perinatal heptachlor exposure increases expression of presynaptic dopaminergic markers in mouse striatum. *Neurotoxicology, 26*(4), 721-728. doi:10.1016/j.neuro.2004.09.003

Chamberlain, C. E., Jeong, J., Guo, C., Allen, B. L., & McMahon, A. P. (2008). Notochord-derived Shh concentrates in close association with the apically positioned basal body in neural target cells and forms a dynamic gradient during neural patterning. *Development, 135*(6), 1097-1106. doi:10.1242/dev.013086

Chi, C. L., Martinez, S., Wurst, W., & Martin, G. R. (2003). The isthmic organizer signal FGF8 is required for cell survival in the prospective midbrain and cerebellum. *Development, 130*(12), 2633-2644.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., & Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature, 383*(6599), 407-413.

Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M., & Peterson, C. (2006). Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol, 2*(9), e123. doi:10.1371/journal.pcbi.0020123

Cole, T. B., Beyer, R. P., Bammler, T. K., Park, S. S., Farin, F. M., Costa, L. G., & Furlong, C. E. (2011). Repeated developmental exposure of mice to chlorpyrifos oxon is associated with paraoxonase 1 (PON1)-modulated effects on cerebellar gene expression. *Toxicological Sciences, 123*(1), 155-169.

Deng, Q., Andersson, E., Hedlund, E., Alekseenko, Z., Coppola, E., Panman, L., . . . Perlmann, T. (2011). Specific and integrated roles of Lmx1a, Lmx1b and Phox2a in ventral midbrain development. *Development, 138*(16), 3399-3408. doi:10.1242/dev.065482

Di Martino, A., Kelly, C., Grzadzinski, R., Zuo, X. N., Mennes, M., Mairena, M. A., . . . Milham, M. P. (2011). Aberrant striatal functional connectivity in children with autism. *Biol Psychiatry, 69*(9), 847-856. doi:10.1016/j.biopsych.2010.10.029

Duff, C., Smith-Miles, K., Lopes, L., & Tian, T. (2012). Mathematical modelling of stem cell differentiation: the PU.1-GATA-1 interaction. *J Math Biol, 64*(3), 449-468. doi:10.1007/s00285-011-0419-3

Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P., Polioudakis, D., . . . Zeng, H. (2017). The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. *Neuron, 96*(3), 542-557. doi:10.1016/j.neuron.2017.10.007

Elwan, M. A., Richardson, J. R., Guillot, T. S., Caudle, W. M., & Miller, G. W. (2006). Pyrethroid pesticide-induced alterations in dopamine transporter function. *Toxicol Appl Pharmacol, 211*(3), 188-197. doi:10.1016/j.taap.2005.06.003

Enver, T., Pera, M., Peterson, C., & Andrews, P. W. (2009). Stem cell states, fates, and the rules of attraction. *Cell Stem Cell, 4*(5), 387-397. doi:10.1016/j.stem.2009.04.011

Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience, 17*(8), 524.

Grandjean, P., & Landrigan, P. J. (2014). Neurobehavioural effects of developmental toxicity. *The lancet neurology, 13*(3), 330-338.

Greene, D. J., Colich, N., Iacoboni, M., Zaidel, E., Bookheimer, S. Y., & Dapretto, M. (2011). Atypical neural networks for social orienting in autism spectrum disorders. *Neuroimage, 56*(1), 354-362. doi:10.1016/j.neuroimage.2011.02.031

Hadley, J. A., Nenert, R., Kraguljac, N. V., Bolding, M. S., White, D. M., Skidmore, F. M., . . . Lahti, A. C. (2014). Ventral tegmental area/midbrain functional connectivity and response to antipsychotic medication in schizophrenia. *Neuropsychopharmacology, 39*(4), 1020-1030.

Harris, R. E., Pargett, M., Sutcliffe, C., Umulis, D., & Ashe, H. L. (2011). Brat promotes stem cell differentiation via control of a bistable switch that restricts BMP signaling. *Dev Cell, 20*(1), 72-83. doi:10.1016/j.devcel.2010.11.019

Hatcher, J. M., Richardson, J. R., Guillot, T. S., McCormack, A. L., Di Monte, D. A., Jones, D. P., . . . Miller, G. W. (2007). Dieldrin exposure induces oxidative damage in the mouse nigrostriatal dopamine system. *Exp Neurol, 204*(2), 619-630. doi:10.1016/j.expneurol.2006.12.020

Hegarty, S. V., Sullivan, A. M., & O'Keeffe, G. W. (2013). Midbrain dopaminergic neurons: a review of the molecular circuitry that regulates their development. *Dev Biol, 379*(2), 123-138. doi:10.1016/j.ydbio.2013.04.014

Hegarty, S. V., Sullivan, A. M., & O'keeffe, G. W. (2013). Midbrain dopaminergic neurons: a review of the molecular circuitry that regulates their development. *Developmental biology, 379*(2), 123-138.

Hook, P. W., McClymont, S. A., Cannon, G. H., Law, W. D., Morton, A. J., Goff, L. A., & McCallion, A. S. (2018). Single-Cell RNA-Seq of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease. *Am J Hum Genet, 102*(3), 427-446. doi:10.1016/j.ajhg.2018.02.001

Horowski, R., Horowski, L., Phil, C., Vogel, S., Poewe, W., & Kielhorn, F.-W. (1995). An essay on Wilhelm von Humboldt and the shaking palsy: first comprehensive description of Parkinson's disease by a patient. *Neurology, 45*(3), 565-568.

Huang, S., Eichler, G., Bar-Yam, Y., & Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett, 94*(12), 128701. doi:10.1103/PhysRevLett.94.128701

Joyner, A. L., Liu, A., & Millet, S. (2000). Otx2, Gbx2 and Fgf8 interact to position and maintain a mid-hindbrain organizer. *Curr Opin Cell Biol, 12*(6), 736-741. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/11063941

Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet, 6*(6), 451-464. doi:10.1038/nrg1615

Kele, J., Simplicio, N., Ferri, A. L., Mira, H., Guillemot, F., Arenas, E., & Ang, S. L. (2006). Neurogenin 2 is required for the development of ventral midbrain dopaminergic neurons. *Development, 133*(3), 495-505. doi:10.1242/dev.02223

Kittappa, R., Chang, W. W., Awatramani, R. B., & McKay, R. D. (2007). The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age. *PLoS Biol, 5*(12), e325. doi:10.1371/journal.pbio.0050325

Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nature communications, 10*(1), 1-14.

Kohls, G., Schulte-Ruther, M., Nehrkorn, B., Muller, K., Fink, G. R., Kamp-Becker, I., . . . Konrad, K. (2013). Reward system dysfunction in autism spectrum disorders. *Soc Cogn Affect Neurosci, 8*(5), 565-572. doi:10.1093/scan/nss033

Kohls, G., Yerys, B. E., & Schultz, R. T. (2014). Striatal development in autism: repetitive behaviors and the reward circuitry. *Biol Psychiatry, 76*(5), 358-359. doi:10.1016/j.biopsych.2014.07.010

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., . . . Linnarsson, S. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell, 167*(2), 566-580 e519. doi:10.1016/j.cell.2016.09.027

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., . . . Furlan, A. (2018). RNA velocity of single cells. *Nature, 560*(7719), 494-498.

Langen, M., Schnack, H. G., Nederveen, H., Bos, D., Lahuis, B. E., de Jonge, M. V., . . . Durston, S. (2009). Changes in the developmental trajectories of striatum in autism. *Biol Psychiatry, 66*(4), 327-333. doi:10.1016/j.biopsych.2009.03.017

Martinat, C., Bacci, J. J., Leete, T., Kim, J., Vanti, W. B., Newman, A. H., . . . Abeliovich, A. (2006). Cooperative transcription activation by Nurr1 and Pitx3 induces embryonic stem cell maturation to the midbrain dopamine neuron phenotype. *Proc Natl Acad Sci U S A, 103*(8), 2874-2879. doi:10.1073/pnas.0511153103

Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M. S., Ko, S. B., Gouda, N., . . . Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics, 33*(15), 2314-2321.

Mavromatakis, Y. E., Lin, W., Metzakopian, E., Ferri, A. L., Yan, C. H., Sasaki, H., . . . Ang, S. L. (2011). Foxa1 and Foxa2 positively and negatively regulate Shh signalling to specify ventral midbrain progenitor identity. *Mech Dev, 128*(1-2), 90-103. doi:10.1016/j.mod.2010.11.002

Meyers, E. N., Lewandoski, M., & Martin, G. R. (1998). An Fgf8 mutant allelic series generated by Cre-and Flp-mediated recombination. *Nature genetics, 18*(2), 136-141.

Mojtahedi, M., Skupin, A., Zhou, J., Castano, I. G., Leong-Quong, R. Y., Chang, H., . . . Huang, S. (2016). Cell Fate Decision as High-Dimensional Critical State Transition. *PLoS Biol, 14*(12), e2000640. doi:10.1371/journal.pbio.2000640

Moreira, E. G., Yu, X., Robinson, J. F., Griffith, W., Hong, S. W., Beyer, R. P., . . . Faustman, E. M. (2010). Toxicogenomic profiling in maternal and fetal rodent brains following gestational exposure to chlorpyrifos. *Toxicology and applied pharmacology, 245*(3), 310-325.

Nakamura, K., Sekine, Y., Ouchi, Y., Tsujii, M., Yoshikawa, E., Futatsubashi, M., . . . Mori, N. (2010). Brain serotonin and dopamine transporter bindings in adults with high-functioning autism. *Arch Gen Psychiatry, 67*(1), 59-68. doi:67/1/59 [pii]

10.1001/archgenpsychiatry.2009.137

Nunes, I., Tovmasian, L. T., Silva, R. M., Burke, R. E., & Goff, S. P. (2003). Pitx3 is required for development of substantia nigra dopaminergic neurons. *Proc Natl Acad Sci U S A, 100*(7), 4245-4250. doi:10.1073/pnas.0230529100

Poulin, J. F., Zou, J., Drouin-Ouellet, J., Kim, K. Y., Cicchetti, F., & Awatramani, R. B. (2014). Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. *Cell Rep, 9*(3), 930-943. doi:10.1016/j.celrep.2014.10.008

Raj, A., & van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell, 135*(2), 216-226. doi:10.1016/j.cell.2008.09.050

Richardson, J. R., Caudle, W. M., Wang, M., Dean, E. D., Pennell, K. D., & Miller, G. W. (2006). Developmental exposure to the pesticide dieldrin alters the dopamine system and increases neurotoxicity in an animal model of Parkinson's disease. *FASEB J, 20*(10), 1695-1697. doi:10.1096/fj.06-5864fje

Richardson, J. R., Caudle, W. M., Wang, M. Z., Dean, E. D., Pennell, K. D., & Miller, G. W. (2008). Developmental heptachlor exposure increases susceptibility of dopamine neurons to N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)in a gender-specific manner. *Neurotoxicology, 29*(5), 855-863. doi:10.1016/j.neuro.2008.05.007

Richardson, J. R., Taylor, M. M., Shalat, S. L., Guillot III, T. S., Caudle, W. M., Hossain, M. M., . . . Miller, G. W. (2015). Developmental pesticide exposure reproduces features of attention deficit hyperactivity disorder. *The FASEB Journal, 29*(5), 1960-1972.

Richardson, J. R., Taylor, M. M., Shalat, S. L., Guillot, T. S., 3rd, Caudle, W. M., Hossain, M. M., . . . Miller, G. W. (2015). Developmental pesticide exposure reproduces features of

attention deficit hyperactivity disorder. *FASEB J, 29*(5), 1960-1972. doi:10.1096/fj.14-260901

Roberts, E. M., English, P. B., Grether, J. K., Windham, G. C., Somberg, L., & Wolff, C. (2007). Maternal residence near agricultural pesticide applications and autism spectrum disorders among children in the California Central Valley. *Environ Health Perspect, 115*(10), 1482-1489. doi:10.1289/ehp.10168

Roberts, J. R., & Karr, C. J. (2012). Pesticide exposure in children. *Pediatrics, 130*(6), e1765-e1788.

Sagner, A., Gaber, Z. B., Delile, J., Kong, J. H., Rousso, D. L., Pearson, C. A., . . . Novitch, B. G. (2018). Olig2 and Hes regulatory dynamics during motor neuron differentiation revealed by single cell transcriptomics. *PLoS Biol, 16*(2), e2003127. doi:10.1371/journal.pbio.2003127

Scott-Van Zeeland, A. A., Abrahams, B. S., Alvarez-Retuerto, A. I., Sonnenblick, L. I., Rudie, J. D., Ghahremani, D., . . . Bookheimer, S. Y. (2010). Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene CNTNAP2. *Sci Transl Med, 2*(56), 56ra80. doi:10.1126/scitranslmed.3001344

Slotkin, T. A., Wrench, N., Ryde, I. T., Lassiter, T. L., Levin, E. D., & Seidler, F. J. (2009). Neonatal parathion exposure disrupts serotonin and dopamine synaptic function in rat brain regions: modulation by a high-fat diet in adulthood. *Neurotoxicol Teratol, 31*(6), 390-399. doi:10.1016/j.ntt.2009.07.003

Smidt, M. P. (2017). Molecular Programming of Mesodiencephalic Dopaminergic Neuronal Subsets. *Front Neuroanat, 11*, 59. doi:10.3389/fnana.2017.00059

Smits, S. M., Burbach, J. P., & Smidt, M. P. (2006). Developmental origin and fate of meso-diencephalic dopamine neurons. *Prog Neurobiol, 78*(1), 1-16. doi:10.1016/j.pneurobio.2005.12.003

Stern, C. (1986). Handbook of Chemical Neuroanatomy: Methods in Chemical Neuroanatomy. Edited by A. Bjorklund and T. Hokfelt. Elsevier, Amsterdam, 1983. Cloth bound, 548 pp. UK£ 140.(Volume 1 in the series). In: Pergamon.

Stott, S. R., Metzakopian, E., Lin, W., Kaestner, K. H., Hen, R., & Ang, S. L. (2013). Foxa1 and foxa2 are required for the maintenance of dopaminergic properties in ventral midbrain neurons at late embryonic stages. *J Neurosci, 33*(18), 8022-8034. doi:10.1523/JNEUROSCI.4774-12.2013

Symmons, O., & Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Mol Cell, 62*(5), 788-802. doi:10.1016/j.molcel.2016.05.023

Torres-Altoro, M. I., Mathur, B. N., Drerup, J. M., Thomas, R., Lovinger, D. M., O'Callaghan, J. P., & Bibb, J. A. (2011). Organophosphates dysregulate dopamine signaling, glutamatergic neurotransmission, and induce neuronal injury markers in striatum. *J Neurochem, 119*(2), 303-313. doi:10.1111/j.1471-4159.2011.07428.x

Vernay, B., Koch, M., Vaccarino, F., Briscoe, J., Simeone, A., Kageyama, R., & Ang, S. L. (2005). Otx2 regulates subtype specification and neurogenesis in the midbrain. *J Neurosci, 25*(19), 4856-4867. doi:10.1523/JNEUROSCI.5158-04.2005

Villaescusa, J. C., Li, B., Toledo, E. M., Rivetti di Val Cervo, P., Yang, S., Stott, S. R., . . . Laguna-Goya, R. (2016). A PBX1 transcriptional network controls dopaminergic neuron development and is impaired in Parkinson's disease. *The EMBO journal, 35*(18), 1963-1978.

Waddington, C. H. (1957). *The strategy of the genes; a discussion of some aspects of theoretical biology*. London,: Allen & Unwin.

Wagner-Schuman, M., Richardson, J. R., Auinger, P., Braun, J. M., Lanphear, B. P., Epstein, J. N., . . . Froehlich, T. E. (2015). Association of pyrethroid pesticide exposure with attention-deficit/hyperactivity disorder in a nationally representative sample of US children. *Environmental Health, 14*(1), 1-9.

Wilson, W. W., Shapiro, L. P., Bradner, J. M., & Caudle, W. M. (2014). Developmental exposure to the organochlorine insecticide endosulfan damages the nigrostriatal dopamine system in male offspring. *Neurotoxicology, 44*, 279-287. doi:10.1016/j.neuro.2014.07.008

Yi, S. H., He, X. B., Rhee, Y. H., Park, C. H., Takizawa, T., Nakashima, K., & Lee, S. H. (2014). Foxa2 acts as a co-activator potentiating expression of the Nurr1-induced DA phenotype via epigenetic regulation. *Development, 141*(4), 761-772. doi:10.1242/dev.095802

Zetterstrom, R. H., Solomin, L., Jansson, L., Hoffer, B. J., Olson, L., & Perlmann, T. (1997). Dopamine neuron agenesis in Nurr1-deficient mice. *Science, 276*(5310), 248-250. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9092472

Zhang, Q., Bhattacharya, B., & Andersen, M. E. (2010, May 24-25, 2010). *A Stochastic Model of the Transcriptional Network Motifs Regulating Hierarchical Cell Lineage Specification and Cellular Reprogramming.* Paper presented at the Conference on Systems Biology of Stem Cells, Irvine, CA.

Zhang, Q., Bhattacharya, S., Kline, D. E., Crawford, R. B., Conolly, R. B., Thomas, R. S., . . . Andersen, M. E. (2010). Stochastic modeling of B lymphocyte terminal differentiation and its suppression by dioxin. *BMC Syst Biol, 4*, 40. doi:10.1186/1752-0509-4-40

Zhang, Q., Kline, D. E., Bhattacharya, S., Crawford, R. B., Conolly, R. B., Thomas, R. S., . . . Kaminski, N. E. (2013). All-or-none suppression of B cell terminal differentiation by environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Appl Pharmacol, 268*(1), 17-26. doi:10.1016/j.taap.2013.01.015

Figures

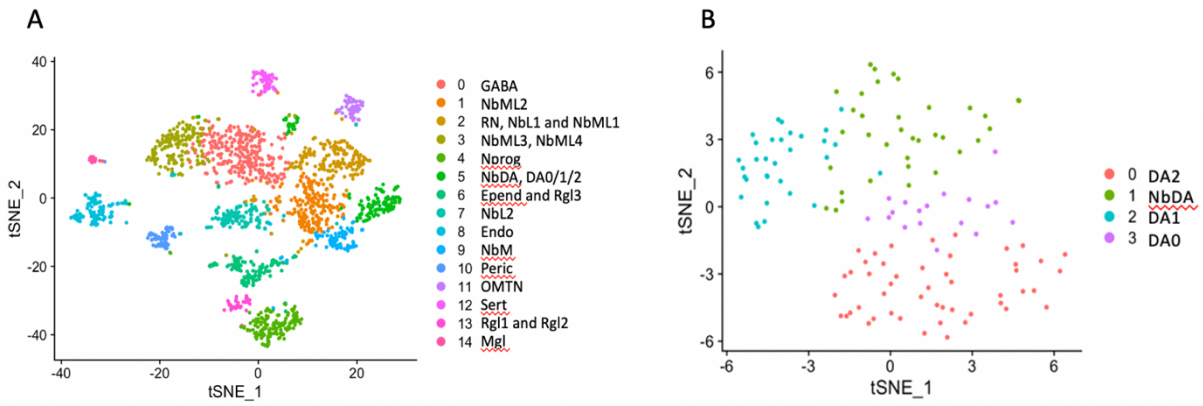

Figure 1. A: 2D t-SNE dimension reduction of mouse midbrain scRNA-seq data from La Manno et al (G. La Manno et al., 2016), showing separation of major cell types and clustering of similar cell types by using Seurat package in R. B: 2D t-SNE dimension reduction of sub-clustering immature DA neurons. Each dot is one cell.

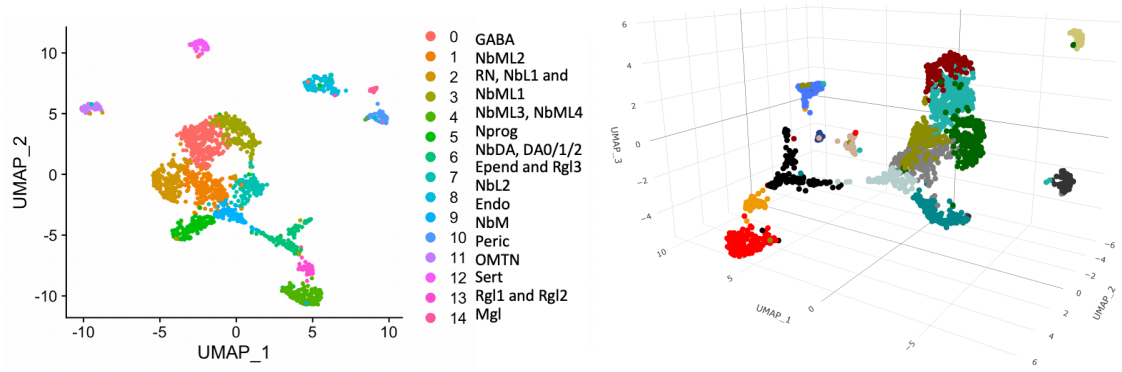

Figure 2. 2D and 3D UMAP dimension reduction of mouse midbrain scRNA-seq data from La Manno et al (G. La Manno et al., 2016), showing separation of major cell types and clustering of similar cell types by using Seurat package in R. Each dot is one cell.

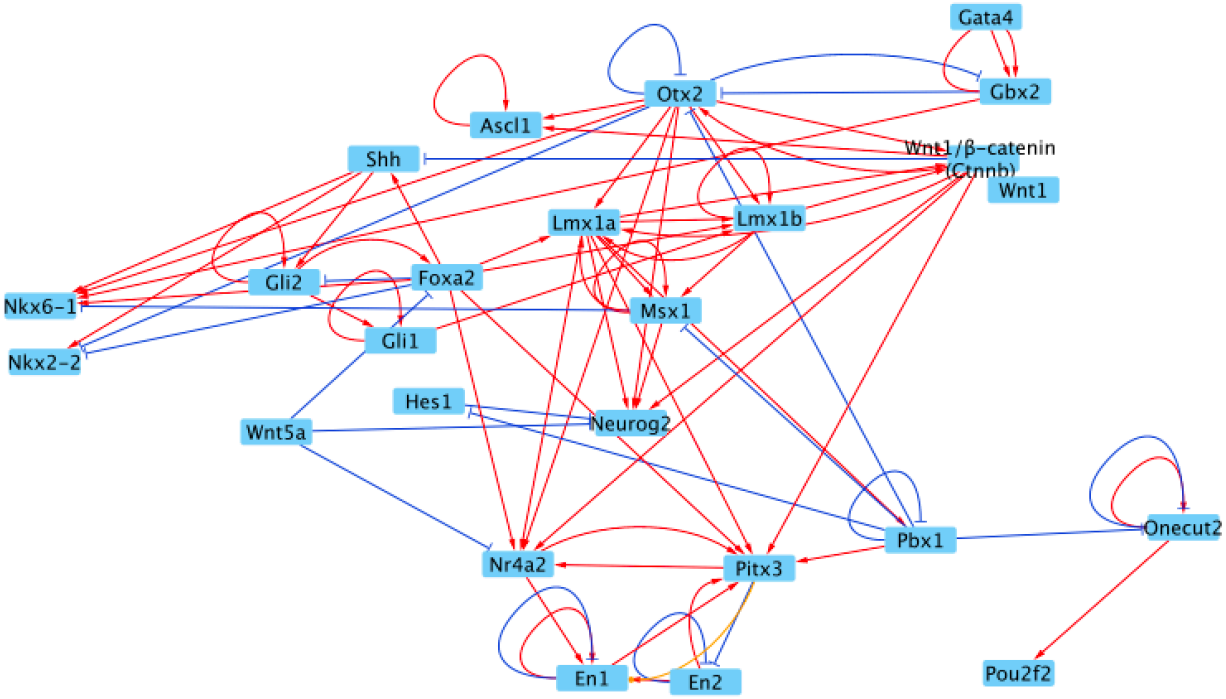

Figure 3. Gene Regulatory Network obtained from pySCENIC and visualized in Cytoscape. Each node is a transcription factor. Each edge is pointing from regulator (transcription factor) to target (genes). Edge with solid target arrow in red: activation. Edge with blunted target arrow in blue: inhibition. Edge in orange means the regulatory direction from pySCENIC's output conflicts with the literature.

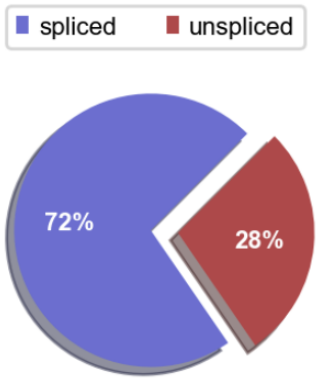

Figure 4. Relative abundance of unspliced over spliced RNA of mouse midbrain scRNA-seq data from La Manno et al (G. La Manno et al., 2016).

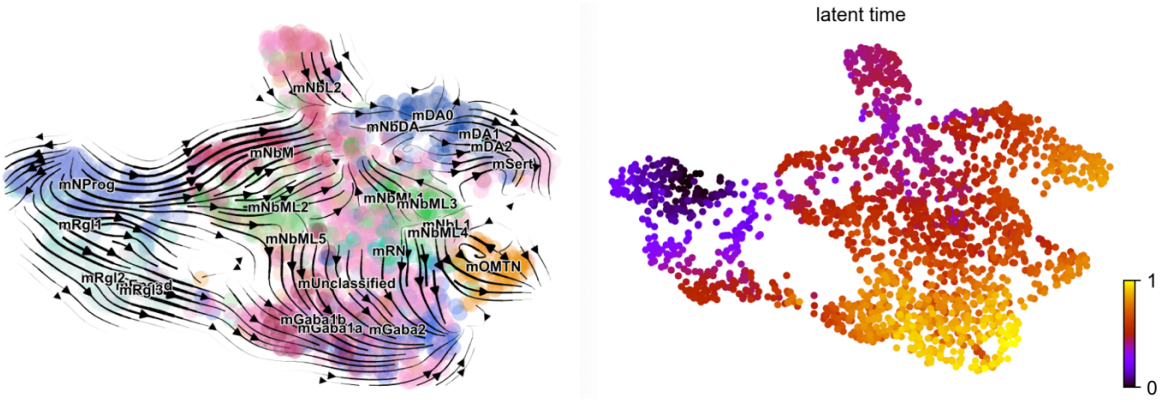

Figure 5. RNA velocity and latent time analysis of mouse midbrain scRNA-seq data from La Manno et al (G. La Manno et al., 2016) projected to 2D UMAP map.

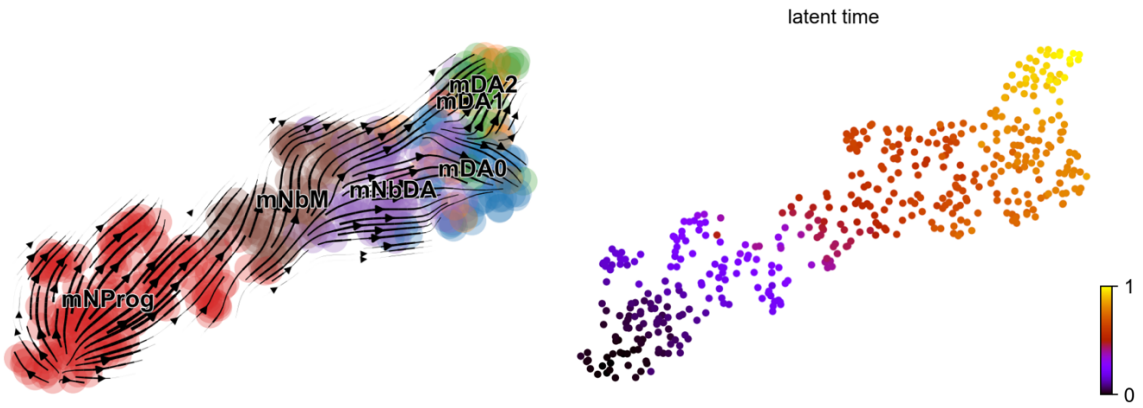

Figure 6. RNA velocity and latent time analysis of DA lineage neurons projected to 2D UMAP map.

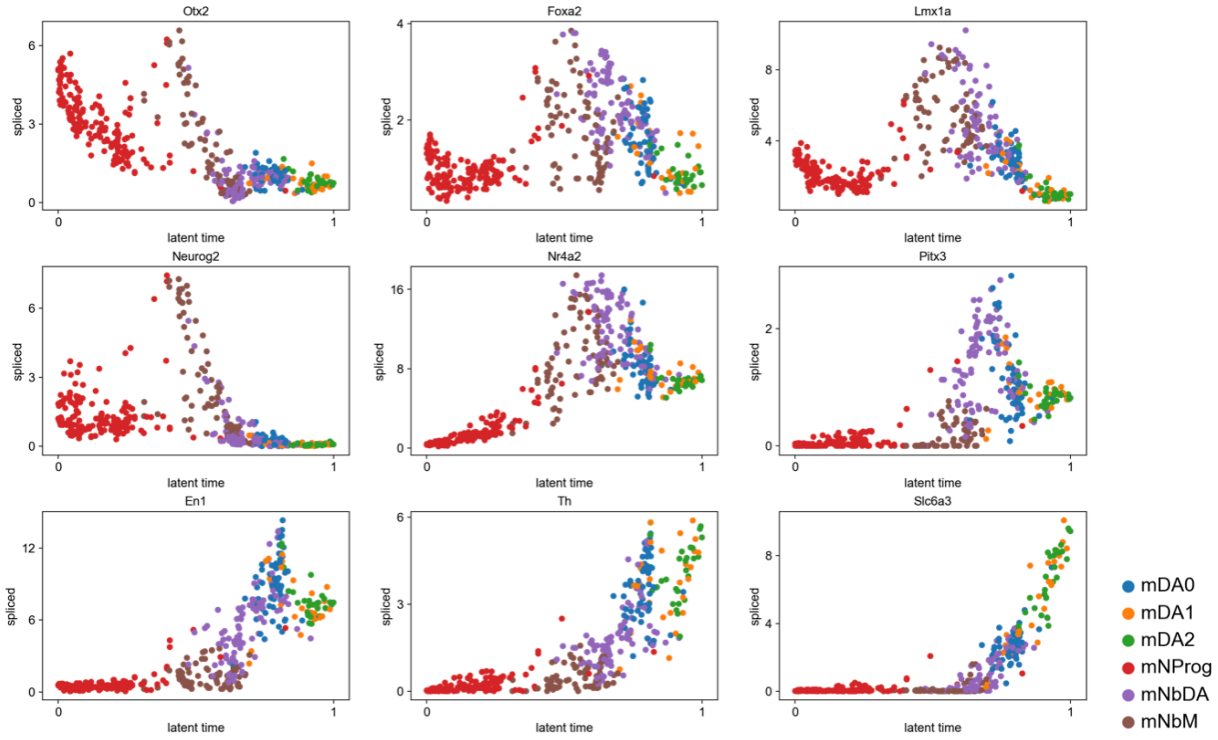

Figure 7. The expression of major marker genes for DA neurons over the predicted latent time.

Tables

Table 1. Marker genes for midbrain neural cells.

Cell Type	**Marker Genes**
Endo	Cldn5, Kdr
Peric	Cd248, Rgs5, Lama4, Gng11
Mgl	Ccl4, Cd53
Epend	Ccdc153, Ttc29, Spag17, Enkur, Ccdc135, Wdr52, Ccdc114, Rfx4, Vit, Ednrb, Celsr1
NbDA	Ebf2, Pitx3, En1, Bnc2
NbM	Neurod1, Igfbpl1, Neurod2, Nhlh1
NProg	Msx2, Atf7ip2, Wnt8b, Neurog1, Hmgn2, Gpsm2, Cenpf, Pbk
Rgl1	Stbd1, Insl6, Hmgn2, Gpsm2, Cenpf, Pbk
Rgl2	Slc6a11, Aldoc, Tnc, Pbk
Rgl3	Cd36, Myh8, Dcn, Vgll3, Sulf1, Cp, Ntn1, Enkur, Ccdc135, Wdr52, Ccdc114
DA2	Aldh1a1
mDA0/1/2	Th, Ajap1, Slc6a3, Syt17, Lmo3
NbML1	Cartpt, Nkx6-2
NbML2	Tal2, Lhx1
NbML3	2610037D02Rik, Calb2, Ldb2
NbML4	Meis2, Grb14, Pnoc
NbL1	Shox2, Crabp1, Lhx9, Lhx2, Tcf7l2, Onecut2, Onecut3
NbL2	Tcp10b, Six4, Cd226, Tcf7l2, Onecut2, Onecut3

RN	D130009I18Rik, Krt20, Pou4f1, Cnih3
OMTN	Isl1, Ecel1, Nefm, Phox2b, Phox2a, Chat, Slc18a3, Lhx4, Tbx20
Sert	Tph2, Fev, Calcr, Hs6st2
Gaba1a	Sst, Gad1, Pnoc
Gaba1b	3110079O15Rik, Gad1, Pnoc
Gaba2	Vsx2, Cdhr1, Npy