Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Prog Biophys Mol Biol. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Prog Biophys Mol Biol. 2015 January ; 117(1): 87–98. doi:10.1016/j.pbiomolbio.2014.11.004.

Identifying Functional Gene Regulatory Network Phenotypes Underlying Single Cell Transcriptional Variability

James Park[1,2], **Babatunde Ogunnaike**[1], **James Schwaber**[1,2], and **Rajanikanth Vadigepalli**[1,2,*]

[1]Department of Chemical and Biochemical Engineering, University of Delaware, Newark, DE 19716

[2]Daniel Baugh Institute for Functional Genomics and Computational Biology, Department of Pathology, Anatomy and Cell Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA 19107

Summary/abstract

Recent analysis of single-cell transcriptomic data has revealed a surprising organization of the transcriptional variability pervasive across individual neurons. In response to distinct combinations of synaptic input-type, a new organization of neuronal subtypes emerged based on transcriptional states that were aligned along a gradient of correlated gene expression. Individual neurons traverse across these transcriptional states in response to cellular inputs. However, the regulatory network interactions driving these changes remain unclear. Here we present a novel fuzzy logic-based approach to infer quantitative gene regulatory network models from highly variable single-cell gene expression data. Our approach involves developing an *a priori* regulatory network that is then trained against *in vivo* single-cell gene expression data in order to identify causal gene interactions and corresponding quantitative model parameters. Simulations of the inferred gene regulatory network response to experimentally observed stimuli levels mirrored the pattern and quantitative range of gene expression across individual neurons remarkably well. In addition, the network identification results revealed that distinct regulatory interactions, coupled with differences in the regulatory network stimuli, drive the variable gene expression patterns observed across the neuronal subtypes. We also identified a key difference between the neuronal subtype-specific networks with respect to negative feedback regulation, with the catecholaminergic subtype network lacking such interactions. Furthermore, by varying regulatory network stimuli over a wide range, we identified several cases in which divergent neuronal subtypes could be driven towards similar transcriptional states by distinct stimuli operating on subtype-specific regulatory networks. Based on these results, we conclude that heterogeneous single-cell gene expression profiles should be interpreted through a regulatory network modeling

© 2014 Elsevier Ltd. All rights reserved

*Correspondence to: Rajanikanth.Vadigepalli@jefferson.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest

The authors declare no conflict of interest

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

perspective in order to separate the contributions of network interactions from those of cellular inputs.

1. Introduction

We recently reported that the variability observed in the transcriptional states of single brainstem neurons can be understood in terms of the distinct combinatorial synaptic inputs each neuron receives (Park, Brureau et al. 2014). These inputs drive individual neurons into distinct neuronal subtypes that lie along a transcriptional landscape characterized by a gene expression gradient. Based on these results, we hypothesized that these emergent neuronal subtypes reflect distinct gene regulatory networks underlying the transcriptional states of individual neurons. There is a need, however, for a robust approach to derive data-driven causal network hypotheses that can be used to interpret and predict the transcriptional behavior of single cells along this transcriptional landscape.

Inferring underlying gene regulatory networks via statistical analysis of single-cell transcription is often complicated by extensive single-cell heterogeneity. However, information about underlying regulatory networks are often manifest in the form of correlations observed in gene expression patterns across single cells. Consequently, single-cell transcriptomic data sets provide a rich experimental sampling of transcriptional states over a wide range of cellular response that can then be used to infer the underlying regulatory network structure (Guo et al. 2010; Buganim et al. 2012a; Janes et al. 2010; Junker & van Oudenaarden 2014). Several methods have been previously developed for deducing regulatory network structures from gene expression data. Statistically-based approaches rely on correlational relationships and dependencies to cluster gene expression profiles, with the rationale being that co-expressed genes are likely to be functionally related (Butte et al. 2000; Zhang & Horvath 2005). One concern with these methods is that the correlational relationships confound direct and indirect effects and do not necessarily imply causal interactions.

Other approaches such as ARACNE overcome these limitations by employing information-theoretic approaches to distinguish between direct and indirect gene interactions (Margolin et al. 2006). Alternatively, Boolean and Bayesian networks have been used successfully to identify regulatory interactions. Although Boolean models characterize genes in a simplified binary ON-OFF state, large-scale computable network models can be generated and analyzed for insights into signaling pathways and biological function (Saez-Rodriguez et al. 2009; Bulashevska & Eils 2005). Bayesian network models provide a probabilistic framework that integrates gene expression data, for example, with *a priori* knowledge of the biological system. While Bayesian network models typically discretize expression data as well, though not necessarily in a binary manner, this approach has been successfully employed to identify gene interaction networks associated with cell function regulation and disease states (Friedman et al. 2000; Kunkle et al. 2013; Friedman 2004).

To overcome some of the limitations of the discretization-based methods, fuzzy logic approaches that consider biologically relevant continuum of gene expression have been used to develop regulatory network models (Woolf et al. 2000; Brock et al. 2009; Sokhansanj et

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

al. 2004; Wang et al. 2002; Zhang et al. 2007). In this study, we develop a novel approach that builds upon the principles and prior applications of fuzzy logic (Tong 1977; Zadeh 1965; Seng et al. 1999; Ying 2003), for analyzing high-throughput single-cell transcriptomic data to identify and model regulatory networks quantitatively. Using this approach, we identified and analyzed the regulatory networks underlying the structured variability of gene expression corresponding to multiple neuronal subtypes (Figure 1A) (Park, Brureau et al. 2014).

2. Methods

2.1. Preliminaries of fuzzy logic

Fuzzy logic is a form of multi-valued logic that deals with approximate reasoning through the use of fuzzy sets (Zadeh 1965). Fuzzy sets contain a continuum of graded membership, ranging from 0 to 1, which allows a variable to transition gradually across multiple membership classes. Fuzzy logic provides an appealing method for describing complex systems using a set of linguistic rules, easily interpretable and expandable, derived from expert knowledge. The ability to handle uncertain, vague, or incomplete information makes fuzzy logic modeling a viable approach for modeling complex biological systems such as gene regulatory networks, as prior work has successfully shown (Woolf et al. 2000; Sokhansanj et al. 2004; Linden & Bhaya 2007; H. Ressom et al. 2003).

A scalability issue, however, arises as the number of process variables and states increases: the proliferation of rules required to explore all possible combinatorial relationships results in computationally intractable models. Several strategies have been employed to address this issue. For example, prior work has demonstrated that constraining the number of rules produce accurate regulatory network models (Sokhansanj et al. 2004). Optimization techniques such as genetic algorithms have been employed to derive an optimal number of rules (Jin 2000; Jin et al. 2008; Linden & Bhaya 2007). The genetic algorithm utilizes the principles of natural evolution in order to seek solutions in a large or possibly infinitely large search space (Martínez-Ballesteros et al. 2014). More recently, Morris et al. (2011) integrated these strategies into a “constrained fuzzy logic” (cFL) approach that allowed them to develop quantitative protein signaling networks, which they used to investigate protein-signaling network behaviors in response to various perturbations. A limitation for utilizing cFL to infer gene regulatory networks is that the approach was formulated for data from experiments that maximally stimulated or inhibited pathways via saturating doses of a ligand or drug. Such targeted, binary-type perturbations do not mimic the continuous fluctuations and variability observed in single-cell transcriptomic data sets (Park, Brureau et al. 2014; Trapnell et al. 2010).

2.2. Adapting constrained Fuzzy Logic

In this study, we formulated a novel cFL approach that is applicable to single cell transcriptomic data. Using this method, we develop an *a priori* gene regulatory network and train it against *in vivo* gene expression measures from single brainstem neurons (Figure 1B). The *a priori* gene regulatory network consists of the direct and indirect causal interactions curated from 1) literature, 2) transcription factor databases, and 3) gene pairs with high

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

correlation in the single neuron expression profiles. We were interested in exploring multiple gene regulatory pathways and therefore included a large number of potential gene interactions. Similar to the earlier cFL approach of Morris et al. (2011), each interaction within the network was modeled as an input-output reaction, where an “input-output” pairing refers to gene nodes connected by a directed edge, mathematically defined by a transfer function. In order to model biologically relevant nonlinear gene interactions, transfer functions approximating Hill functions were used (Figure 2B). To address reactions involving multiple gene inputs and facilitate development of a computable model, the fuzzy logic framework relies on Zadeh fuzzy logic gates (Figure 2B) (Zadeh 1965; Jin & Wang 2009) to determine the output of multi-gene interactions. The network identification approach utilizes a limited number of Hill-function parameter sets spanning a wide range of linear to near-binary "ON-OFF"-type relationships to define quantitatively a particular interaction. The specific Hill-function parameter values are identified based on the experimental data in a subsequent training step.

Once the *a priori* gene regulatory network is specified, it is then trained against the experimental transcriptional data set via a genetic algorithm to determine an optimal network structure capable of fitting the experimental data below a predetermined error threshold. The mean square error (MSE) was used as a measure of the fit and predictive capability of the network model. Following the optimization scheme of Morris et al. (2011), each independent run of the genetic algorithm generated a population of regulatory network models (referred to as optimized *unprocessed* models), with the “best” network model (i.e., lowest MSE) selected for further refinement. The Hill-function parameters in the selected *unprocessed* model were then further refined using a non-linear optimization scheme based on the subplex algorithm (Johnson 2008), which produces an *unprocessed-refined* network model.

Although the training step removes reactions not supported by the data, we observed a number of redundant reactions that were included in the original *a priori* network remained in the *unprocessed* network. In order to simplify the model further, the *unprocessed-refined* network model went through a final model reduction step. Here, the frequency of input selection for each gene interaction (as defined by the Zadeh gates–Figure 2B) was determined from simulations performed during the training step. These calculated frequencies were translated into edge weights that indicated the relative “dominance” of a particular gene interaction. In certain cases, some redundant gene interactions were not pruned by the optimization scheme during the model training. These non-dominant or unused gene interactions were removed in order to generate a final *unprocessed-refined-reduced* gene regulatory network model.

Due to the stochastic nature of the genetic algorithm, multiple plausible network solutions can be generated that fit the data equally well, each representing a hypothesized network of regulatory interactions corresponding to the experimentally measured transcriptional states of single cells. We considered an ensemble of network model solutions, generated through multiple iterations of the training, refinement, and reduction processes (Sokhansanj et al. 2004; Marbach et al. 2009; Morris et al. 2011). The abundance or lack of a particular gene interaction across individual network models within an ensemble was interpreted as the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

likelihood of the particular relationship driving that specific response (Datta & Sokhansanj 2007; Marbach et al. 2009).

2.3. Statistical analysis and significance of the a priori network

To assess the fit and predictive capabilities of each *unprocessed-refined-reduced* network model, an n-fold cross validation procedure was performed. In this process, two-thirds of the data set was defined as the training subset used to train, refine, and reduce the *a priori* network. The remaining one-third was set aside as a validation subset that was used to evaluate the predictive capabilities of the *unprocessed-refined-reduced* network. Within an iteration of the cross-validation process, the training-refinement-reduction step was repeated five times in order to identify several different plausible models generated by the genetic algorithm. This step was then repeated forty times resulting in a total of 200 potential network-model solutions. Network models with the corresponding lowest twenty-five MSE values were selected to form a network ensemble.

In addition to cross-validating the *a priori* network, we performed a series of tests to determine the statistical significance of the ensemble of regulatory network models. We compared the subtype-specific network ensembles of *unprocessed* network models to *unprocessed* models trained against randomized data or unprocessed network models derived from randomized *a priori* networks. The randomized *a priori* networks were generated either by 1) randomizing directed edge placement throughout the *a priori* network, or 2) generating a randomize network topology. Both types of randomized networks maintained a directed acyclic structure. This constraint was placed on the network randomization process to avoid any feedback interactions, the effects of which would not be observable in a single time point. Thus training a cyclic network against the single time point data set would falsely increase the MSE of the *unprocessed* randomized cyclic network (refer to section 3.1 for greater details). Corresponding p-values for each subtype-specific network ensemble were empirically determined by comparing the *unprocessed* network ensemble MSE values to MSE values calculated from 500 iterations for each test condition. *A priori* network training with random edge removal was also performed to evaluate the dependence of the training process on the *a priori* network structure.

2.4. Simulation and analysis of single-neuronal transcriptional states

To elucidate the structure of the underlying gene regulatory networks corresponding to neuronal subtypes, we trained the *a priori* network against 1) the scaled gene-expression data for each neuronal subtype identified in our prior analysis, and 2) the scaled gene-expression data set spanning all neuronal subtypes (Park, Brureau et al. 2014). Comparing regulatory network model fits of the network ensembles produced from training by either type of data set would provide insight into whether a common regulatory network or distinct regulatory networks underlie the experimentally observed distribution of neuronal transcriptional states.

Next, we investigated the response of the network ensemble to a range of randomly generated regulatory network stimuli levels using Latin hypercube sampling (LHS). We analyzed the simulated single-cell transcriptional states generated from our network

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ensembles using non-metric multidimensional scaling (MDS). We have previously employed MDS to visualize and interpret high-dimensional single-cell gene expression data and compare the transcriptional states of single neurons. Detailed descriptions of these methods were previously outlined in Park, Brureau et al. (2014).

2.5. Single-cell gene expression data

The high-throughput transcriptional data set analyzed was obtained from our previous study of 300 individual brainstem neurons collected from rats exposed to a pharmacologically induced acute hypertensive challenge (Park, Brureau et al. 2014). Gene expression levels were measured in single neuron samples using a high-throughput qPCR platform (BioMark^{TM}, Fluidigm). Quality control and normalization procedures to generate $-\Delta C_t$ are described in detail in Park, Brureau et al. (2014). Prior to performing fuzzy logic modeling we first normalized the $-\Delta C_t$ gene expression values to a [0, 1] scale. Expression levels for each gene were normalized by subtracting the minimum expression value across all single neuron samples, and dividing by the corresponding range of expression levels. The normalization technique has been used in previous fuzzy logic analyses of gene expression data (Woolf et al. 2000; Brock et al. 2009; H. Ressom et al. 2003).

2.6. Computational platforms

Genetic algorithm training and network ensemble simulations were performed in the R statistical software (R Core Team 2013) using various functions with the CellNOptR and CNORfuzzy packages (Saez-Rodriguez et al. 2009; Morris et al. 2011). Several functions within these packages were modified to account for continuous levels of network inputs, applicable to the present single cell transcriptional data set. MDS analysis and Latin hypercube sampling was performed using the *MASS* package (Venables & Ripley 2002) and *lhs* package respectively, in the R statistical software (R Core Team 2013). All regulatory network figures were generated using Cytoscape v2.8.3 (Shannon et al. 2003)

3. Results

3.1 A priori network of the AT1R-mediated pathway

We employed a fuzzy-logic approach to infer a gene-regulatory network involved in the angiotensin type 1 receptor (AT1R)-mediated pathway. AT1R mediates central autonomic control of blood pressure by the octapeptide angiotensin II (AngII) via modulation of NTS neurons. The activation of AT1R (initiated upon binding with AngII) triggers a signaling cascade resulting in an increased activation of transcription factors (TFs) such as ELK1, FOS, and JUN (Mitra et al. 2010; Sumners et al. 2002), as well as an increase binding activity of AP-1, consisting of phosphorylated FOS and JUN proteins. AP-1 binds to target promoters leading to changes in gene expression of tyrosine hydroxylase (*Th*), dopamine β-hydroxylase (*Dbh*), and norepinephrine transporter (*Slc6a2*), all critical to the production and release of the catecholamine norepinephrine. These changes result in neuromodulation of catecholamines as well as enhanced inhibitory GABAergic transmission, both associated with hypertension (Veerasingham & Raizada 2003).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In order to model the AT1R mediated regulatory network, we formulated a four-tiered hierarchically structured *a priori* set of interactions derived from relevant literature (Figure 2). Since the transcriptional changes previously described occur upon AT1R activation, the corresponding gene, *Agt1r* was positioned in the first tier of the hierarchical structure. And while AT1R mediates the effects of AngII, the latter is the product of the enzymatic reaction between the angiotensin-converting enzyme (ACE) and angiotensinogen. Therefore the genes *Ace* and *Agt* were placed alongside *Agt1r* as upstream-signal (i.e. network) inputs. Stimulating these three first-tier network inputs activates the second tier of TFs. These TFs interact with two gene groups referred to as "feedback" and "downstream target" genes to form third and fourth tiers, respectively. Additional genes were added to the TF, feedback, and downstream target tiers as supported by literature and computational analysis of promoter-gene regions using the PAINT software and the TRANSFAC database (Vadigepalli et al. 2003). The nodes in the a priori network and the relevant literature implicating the role for each network node are summarized in Table 1.

The fuzzy-logic approach of input-output influences cannot handle cycles in the regulatory network. To address this issue, the feedback interactions were considered as affecting the target gene expression levels as a functional consequence of signaling regulation. In the directed acyclic graph, direct feedback inhibition on *Agt1r* coming from *Arrb1, Arrb2, Rgs2, and Rgs4* was rearranged to inhibit expression of *Agt1r* target genes (Figure 2A).

3.2 Network ensembles recapitulate single-neuron transcriptional variability

Using the cross-validation procedure, we evaluated the fit and predictive capabilities of the *unprocessed-refined-reduced* models. With the exception of the second order neuronal subtype-specific network ensemble, each subtype-specific ensemble produced lower MSE values than the common, or general, network ensemble (Figure 3A). However, comparing the fit of the general network ensemble to the second order neuronal-specific data subset revealed a higher MSE than the corresponding subtype-specific network ensemble MSE (Figure 3A). The result indicated that the subtype-specific networks better fit the data.

We further assessed the predictive capabilities of these models using heat maps for simple visual interpretation (Figure 3). Comparison of the simulated transcriptional states generated by a representative *unprocessed-refined-reduced* network model and the corresponding experimental data show similar transcriptional behavior across single cells and genes, further corroborated by the low absolute difference between data and simulation (i.e., residuals) (Figure 3B). A similar performance of model simulations to experimental data was observed across all neuronal subtypes modeled. The improved ability of the subtype-specific network ensembles to fit the transcriptional profiles of individual neurons indicates that the previously observed transcriptional gradient (Park, Brureau et al. 2014) is not driven by some common regulatory network, but rather by distinct networks corresponding to these neuronal subtypes.

3.3 Distinct regulatory network topologies define neuronal subtypes

When comparing the ensemble network topologies of the extreme neuronal subtypes (catecholaminergic and second-order subtypes), several distinct network structures

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

distinguished these two extreme neuronal subtypes. For example, the transcriptional states of *Agt* and *Agtr1a* strongly influence the transcriptional output of the catecholaminergic ensemble network, which corresponds with previous expectations of AT1R mediated effects of AngII on catecholaminergic neurons. These *Agt-Agtr1a* influences, however, were reduced in the second-order ensemble network where Ace shows a stronger influence on subsequent transcription factors, as indicated by edge thickness (Figure 3). This shift in influence suggests that *Agtr1a*-mediated effects were more dependent upon the availability of AngII. Alternatively, the prominent influence of Ace in the individual neurons of the second order neuronal-subtype may reflect the direct effects Ace has on intra-neuronal signaling (Lambert, DW, Clarke, N, et al. 2009).

Additionally, the prominent TF interactions (*Atf2, Fosl1, and Jund*) regulating *Th* and *Dbh* expression in the catecholaminergic-specific ensemble differentiated this subtype from the second order neuronal network ensemble (Figure 4). The presence of the AP-1 interactions in the catecholaminergic ensemble and absence in the second-order ensemble matches previous reports of increased AP-1 activation in catecholaminergic neurons (Veerasingham & Raizada 2003). Second-order neurons do not express *Th* and *Dbh*, and hence the interactions corresponding to AP-1 activation of *Th and Dbh* expression were absent in the second-order ensemble. Likewise, an increase in TF stimulatory interactions for *gal*, *Gad1*, and *Tac1* was observed in the second-order ensemble. Second-order neurons may play an inhibitory role in blood pressure regulation (Veerasingham & Raizada 2003). Our modeling results, which predict an increased number of TF interactions in the second-order network ensemble, suggest a causal link between AT1R-mediated pathway activation and neuronal inhibitory transmission.

The signaling feedback interactions were the distinguishing features of the network ensembles; second-order neuron networks showed evidence for *Rgs2* and *Rgs4* mediated suppression of downstream target genes, such as *Gad1*, unlike the catecholaminergic ensemble. These distinct interactions mediating the effects of *Agt*, *Agtr1a*, and *Ace* in catecholaminergic vs. second-order neurons indicate that these subtypes are driven by distinct regulatory networks. Consistent with these results, the regulatory networks corresponding to the intermediate neuronal subtypes also showed a subset of distinct causal interactions that drive the transcriptional states of individual neurons along a gene expression gradient (Figure 5, Supplemental Figures S1-S4).

3.4 Distinct network topologies and stimuli support a distribution of single-neuron transcriptional states

To investigate how distinct regulatory networks drive the emergent neuronal subtypes, we simulated the transcriptional states of individual neurons by stimulating the network ensembles with the same set of randomly selected stimuli levels of *Agt*, *Agtr1a*, and *Ace* (section 2.4). Using our network ensemble models we reproduced the distribution of neuronal subtypes, as visualized via MDS analysis (Figure 6A). Similar to our previous analysis of experimental data (Park, Brureau et al. 2014), our analysis of simulated transcriptional states yielded distinct clusters of neurons segregated by the regulatory network subtype (Figure 1A and Figure 6A). The differences between transcriptional states

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are driven by subtype-specific regulatory networks receiving similar overall spectrum of inputs.

Within each subtype, we observed a wide distribution of neuronal transcriptional states indicated by the large area covered by each subgroup in the MDS space (Figure 6). In certain cases, the inter-subtype pairwise distance is shorter than the intra-subtype pairwise distance, with the shorter distance between neuronal pairs indicating greater transcriptional similarity despite the differences in the underlying regulatory networks. In these instances, distinct network inputs appear to drive the differing regulatory networks towards a similar transcriptional state (Figure 6). Thus, our fuzzy-logic network modeling predicts that the pervasive transcriptional variability observed *in vivo* is likely a product of distinct regulatory network interactions as well as response to distinct network stimuli (Raj & van Oudenaarden 2008; Eberwine & Bartfai 2011; Kim et al. 2011; Bendall et al. 2012; Pe'er & Hacohen 2011; Amir et al. 2013; Buganim et al. 2012b).

3.5 Statistical significance and predictive capabilities of network ensembles

To further test and verify the predictive capabilities of the fuzzy logic-based models, we evaluated how well a particular subtype-specific ensemble would predict the transcriptional states of catecholaminergic neurons captured and measured independently from the data set used to cross-validate the *a priori* regulatory network. Simulations of the catecholaminergic-specific network ensemble using the scaled network input levels measured from the independent neurons reproduced the experimentally measured transcriptional states remarkably well (Supplement Figure S5). Additionally, statistical analysis of the *unprocessed* network ensembles revealed that nearly all subtype-specific *unprocessed* network models were statistically significant as all other randomized cases generated higher MSE values (empirically determined p values < 0.05). In only one instance did a p-value > 0.05 occur (Th^{low} Fos^{low} network ensemble, p-value = 0.058). While this may not meet the statistical threshold of significance typically used, a p-value of 0.058 indicates a low probability that similar or better network predictions could be obtained from a network trained against randomized data (Supplemental Figure S6).

4. Discussion

In this study, we developed a fuzzy logic modeling approach that supports the inference of causal gene regulatory networks from highly variable *in vivo* single-neuron transcriptomic data. This approach treats the single-neuron transcriptomic data set as a rich source of information about underlying regulatory networks, which manifest as correlated gene expression patterns across hundreds of single neurons. (Park, Brureau et al. 2014). Our quantitative regulatory network models allow us to interpret how a structured organization of neuronal subtypes arises from a distribution of heterogeneous single neurons through regulatory network interactions; distinct regulatory network topologies and their responses to distinct network inputs drive individual neurons through a range of transcriptional states.

In comparison to other fuzzy logic based approaches, we found that our fuzzy logic modeling approach provides a robust technique to infer quantitative regulatory network models from variable single-cell gene expression data. Previous fuzzy logic modeling has

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

identified only qualitative gene regulatory networks from microarray data, in the context of activators, repressors, and target gene triplets (Woolf et al. 2000; Sokhansanj et al. 2004). In instances where quantitative network models have been developed, these models relied on data generated under defined, binary-type experimental perturbations (Morris et al. 2011), which do not capture the continuous changes in network inputs observed *in vivo*. Our fuzzy logic approach, however, accounts for continuous levels of gene regulatory network inputs and allows us to model single-cell gene expression responses to more subtle changes occurring under physiological conditions. This methodology is a novel approach to infer quantitative regulatory network models that provide insight into the regulatory mechanisms contributing to the heterogeneous nature of single cells.

4.1 Distinct regulatory networks underlying heterogeneous responses at the single-cell level

The presence of distinct regulatory structures underlying neuronal variability posits implications as to how a neuronal population will respond to a targeted gene intervention. Apart from the variability one would expect given the stochastic nature of gene expression (Raj & van Oudenaarden 2008), manipulating a specific gene within a neuronal regulatory network is likely to differentially affect neuronal subtypes. As opposed to an “analog” or population response to a targeted gene intervention, our models suggest a more “discrete” or subpopulation-specific response based on their respective and distinct network topology. Our model-based prediction aligns with previous results demonstrating heterogeneous dose response across single cells, as in the case of NF-kB in response to TNFα (Tay et al. 2010). Similarly, heterogeneous responses to drug treatment have been observed in clonal cancer cell lines (Cohen et al. 2008). With multiple studies repeatedly demonstrating a large degree of heterogeneity among individual cells, our network modeling approach provides a path towards understanding the regulatory network topology of cellular subtypes.

Individual cells naturally respond to a variety of environmental cues and stimuli in order to develop properly and perform specific cellular functions. These developmental and functional responses are supported by distinct genetic programs and regulatory network rewiring. While gene regulatory networks are often cast as static snapshots, regulatory networks are cell-, tissue-, and condition-specific and exhibit dynamic adaptation in response to both internal and external signals (Pe’er & Hacohen 2011). For instance, in the context of cancer, regulatory network adaptation is highly prevalent in the differential response between normal and diseased cells. One example involves a drug treatment that targets the ERK pathway. While the drug inhibits the ERK pathway and effectively removes tumors in melanomic patients with an oncogenic BRAF mutation (Flaherty et al. 2010), the same drug has the opposite effect in cells with wild-type BRAF and activates the ERK pathway, potential promoting tumors in those cells (Poulikakos et al. 2010). Moreover, experimental evidence in yeast strains has shown that widespread changes occur in gene interaction networks and the pathways they represent during cellular response to DNA damage (Bandyopadhyay et al. 2010). These reorganized interactions demonstrate that cells rely upon the ability to rewire regulatory network programming in order govern dynamic cellular functions in response to changing environments, stress, and stimuli.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

4.2. Distinct stimuli and regulatory networks drive single cells across a transcriptional landscape

Within the brain, distinct gene regulatory network activity and genetic programs are critical in modulating the transcriptional state of neurons. Specific transcriptional mechanisms shape the developmental trajectory of neurons, which contributes to overall neuronal diversity and connectivity (Chen et al. 2006; Kramer et al. 2006; Friese et al. 2009). Not only are specific genetic programs critical to neuronal development, changes and adaptations of regulatory networks continue to play an essential role in post-mitotic neurons. In response to cellular inputs, distinct regulatory network interactions direct individual neurons into distinct subtypes towards supporting physiological functions. These distinct regulatory networks drive neurons towards specific transcriptional states (Figure 6D). In addition, distinct network inputs stimulating subtype-specific regulatory networks can drive seemingly divergent neuronal subtypes towards similar transcriptional states, with likely impact on physiological function (Figure 6E). Our simulations indicate that the transcriptional states of individual neurons can be tuned via graded inputs to the subtype-specific regulatory networks. The transcriptional (i.e. functional) flexibility and diversity of individual neurons provide a mechanistic explanation to the experimentally observed plasticity of mature neurons (Dulcis et al. 2013; De la Rossa et al. 2013). Furthermore, the tunable nature of the transcriptional states of individual neurons poses intriguing possibilities regarding the dynamics of how neurons traverse across subtype-specific transcriptional states. Similar to developmental trajectory, are there critical regulatory interactions that can shift an individual or a subpopulation of neurons to a different trajectory towards a particular state in the development of disease? Such a question requires time course data from single neurons as well as modifications to our approach to consider dynamics.

4.3. Modeling assumptions and potential extensions of the approach

Our trained network ensembles contain likely gene interactions and influences that fit *in vivo* gene expression data remarkably well. The biological scope of the regulatory interactions and mechanisms predicted by the network ensembles is limited to the data types used to train the network models. The approach could be expanded to include additional regulatory mechanisms such as miRNA interactions, and post-translational regulation. While the mRNA levels of key genes have been shown to correlate with corresponding protein levels and neuronal function (Toledo-Rodriguez et al. 2004; Schulz et al. 2006), this remains an assumption inherent to the hypothesized gene regulatory interactions derived from the experimental data sets from the transcriptional domain. The advantage of fuzzy-logic models lies in their flexibility and capacity to easily expand the *a priori* network to incorporate additional regulatory components and interactions. By adding the appropriate gene-miRNA interaction, for instance, to the *a priori* network, we can extend the biological scope of the regulatory mechanisms involved in gene regulation and overall single-cell heterogeneity. Our study enables new opportunities to employ the fuzzy-logic based network models to inform future experimental design and iteratively refine the network ensembles to further elucidate the mechanisms that support neuronal and other cell-type diversity within the central nervous system.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5. Conclusions

Current high-throughput technologies have provided us with an unprecedented ability to extensively examine single-cell behavior at the molecular level. Accordingly, identifying input-output gene regulatory relationships provides insights into the causal mechanisms that drive cells through a range of phenotypic states. We developed a novel fuzzy-logic-based approach to analyze single-cell *in vivo* transcriptomic data. Using this methodology, we have been able to identify the contributions of regulatory network interactions and those of cellular inputs have in shaping the transcriptional states of neuronal subtypes. Our results support a new approach that goes beyond classifying neurons based on gene expression profiles and analyzes neuronal subtypes based on regulatory network phenotypes, which drive their transcriptional states. Using this approach, we can begin to understand the regulatory mechanisms driving single cell variability that underlie normal and disease physiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The work presented here is funded through NIAAA T32 AA007463, NIH NIGMS R01 GM083108, and NIH NHLBI R01HL111621. J.P. would like to thank Dr. Daniel Lees for his critical review of the paper and Mr. Hiren Makadia for facilitating the computational analyses.

References

Amir ED, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nature biotechnology. 2013; 31(6):545–52.

Bandyopadhyay S, et al. Rewiring of genetic networks in response to DNA damage. Science (New York, N.Y.). 2010; 330(6009):1385–9.

Bendall SC, et al. A deep profiler's guide to cytometry. Trends in immunology. 2012; 33(7):323–32. [PubMed: 22476049]

Brock GN, Beavis WD, Kubatko LS. Fuzzy logic and related methods as a screening tool for detecting gene regulatory networks. Information Fusion. 2009; 10(3):250–259.

Buganim Y, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 2012a; 150(6):1209–22. [PubMed: 22980981]

Buganim Y, et al. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell. 2012b; 150(6):1209–22. [PubMed: 22980981]

Bulashevska S, Eils R. Inferring genetic regulatory logic from expression data. Bioinformatics (Oxford, England). 2005; 21(11):2706–13.

Butte, a J., et al. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97(22):12182–6. [PubMed: 11027309]

Chen C-L, et al. Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. Neuron. 2006; 49(3):365–77. [PubMed: 16446141]

Cohen, a a, et al. Dynamic proteomics of individual cancer cells in response to a drug. Science (New York, N.Y.). 2008; 322(5907):1511–6.

Datta S, Sokhansanj B. a. Accelerated search for biomolecular network models to interpret high-throughput experimental data. BMC bioinformatics. 2007; 8:258. [PubMed: 17640351]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Dulcis D, et al. Neurotransmitter switching in the adult brain regulates behavior. Science (New York, N.Y.). 2013; 340(6131):449–53.

Eberwine J, Bartfai T. Single cell transcriptomics of hypothalamic warm sensitive neurons that control core body temperature and fever response Signaling asymmetry and an extension of chemical neuroanatomy. Pharmacology & therapeutics. 2011; 129(3):241–59. [PubMed: 20970451]

Flaherty KT, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. the new england journal of medicine. 2010; 363(9):809–819. [PubMed: 20818844]

Friedman N. Inferring cellular networks using probabilistic graphical models. Science (New York, N.Y.). 2004; 303(5659):799–805.

Friedman N, et al. Using Bayesian networks to analyze expression data. Journal of computational biology: a journal of computational molecular cell biology. 2000; 7(3-4):601–20. [PubMed: 11108481]

Friese A, et al. Gamma and alpha motor neurons distinguished by expression of transcription factor Err3. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106(32):13588–93. [PubMed: 19651609]

Gasnier B. The SLC32 transporter, a key protein for the synaptic release of inhibitory amino acids. Pflügers Archiv: European journal of physiology. 2004; 447(5):756–9. [PubMed: 12750892]

Guo G, et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Developmental cell. 2010; 18(4):675–85. [PubMed: 20412781]

Hall ME, Miley FB, Stewart JM. Cardiovascular effects of substance P peptides in the nucleus of the solitary tract. Brain Research. 1989; 497(2):280–290. [PubMed: 2479449]

Herdegen T, Leah JD. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. Brain Research Reviews. 1998; 28(3):370–490. [PubMed: 9858769]

Janes, K. a, et al. Identifying single-cell molecular programs by stochastic profiling. Nature methods. 2010; 7(4):311–7. [PubMed: 20228812]

Jin Y. Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement. IEEE Transactions on Fuzzy Systems. 2000; 8(2):212–221.

Jin Y, Member S, Sendhoff B. Evolving in silico Bistable and Oscillatory Dynamics for Gene Regulatory Network Motifs. 2008:386–391.

Jin, Y. Fuzzy Systems in Bioinformatics and Computational Biology. Wang, L., editor. Springer Berlin Heidelberg; Berlin, Heidelberg: 2009.

Johnson SG. The NLopt nonlinear-optimization package. 2008

Junker JP, van Oudenaarden A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. Cell. 2014; 157(1):8–11. [PubMed: 24679522]

Karin M, Liu Z, Zandi E. AP-1 function and regulation. Current Opinion in Cell Biology. 1997; 9(2): 240–246. [PubMed: 9069263]

Kim TK, et al. Transcriptome transfer provides a model for understanding the phenotype of cardiomyocytes. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(29):11918–23. [PubMed: 21730152]

Kramer I, et al. A Role for Runx Transcription Factor Signaling in Dorsal Root Ganglion Sensory Neuron Diversification. Neuron. 2006; 49(3):379–393. [PubMed: 16446142]

Kunkle BW, Yoo C, Roy D. Reverse engineering of modified genes by Bayesian network analysis defines molecular determinants critical to the development of glioblastoma. PloS one. 2013; 8(5):e64140. [PubMed: 23737970]

Kvetnansky R, Sabban EL, Palkovits M. Catecholaminergic Systems in Stress: Structural and Molecular Genetic Approaches. 2009:535–606.

De la Rossa A, et al. In vivo reprogramming of circuit connectivity in postmitotic neocortical neurons. Nature neuroscience. 2013; 16(2):193–200.

Linden R, Bhaya A. Evolving fuzzy rules to model gene expression. Bio Systems. 2007a; 88(1-2):76–91.

Linden R, Bhaya A. Evolving fuzzy rules to model gene expression. Bio Systems. 2007b; 88(1-2):76–91.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Luttrell LM, Lefkowitz RJ. The role of beta-arrestins in the termination and transduction of G-protein-coupled receptor signals. Journal of cell science. 2002; 115:455–65. Pt 3. [PubMed: 11861753]

Marbach D, Mattiussi C, Floreano D. Combining multiple results of a reverse-engineering algorithm: application to the DREAM five-gene network challenge. Annals of the New York Academy of Sciences. 2009a; 1158:102–13. [PubMed: 19348636]

Marbach D, Mattiussi C, Floreano D. Combining multiple results of a reverse-engineering algorithm: application to the DREAM five-gene network challenge. Annals of the New York Academy of Sciences. 2009b; 1158:102–13. [PubMed: 19348636]

Marc Y, Llorens-Cortes C. The role of the brain renin-angiotensin system in hypertension: implications for new treatment. Progress in neurobiology. 2011; 95(2):89–103. [PubMed: 21763394]

Margolin, A. a, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC bioinformatics. 2006; 7(Suppl 1):S7. [PubMed: 16723010]

Martínez-Ballesteros M, Nepomuceno-Chamorro IA, Riquelme JC. Discovering gene association networks by multi-objective evolutionary quantitative association rules. Journal of Computer and System Sciences. 2014; 80(1):118–136.

Mehta PK, Griendling KK. Angiotensin II cell signaling: physiological and pathological effects in the cardiovascular system. 2007:82–97.

Melander T, et al. Coexistence of Galanin-like lmmunoreactivity with Catecholamines , GABA and Neuropeptides in the Rat CNS. 1986:3640–3654. 6(December 1966).

Mitra AK, Gao L, Zucker IH. Angiotensin II-induced upregulation of AT(1) receptor expression: sequential activation of NF-kappaB and Elk-1 in neurons. American journal of physiology. Cell physiology. 2010; 299(3):C561–9. [PubMed: 20554912]

Morris MK, et al. Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. PLoS computational biology. 2011; 7(3):e1001099. [PubMed: 21408212]

Park J, et al. Inputs drive cell phenotype variability. Genome research. 2014; 24(6):930–941. [PubMed: 24671852]

Paton JF, Li YW, Schwaber JS. Response properties of baroreceptive NTS neurons. Annals of the New York Academy of Sciences. 2001; 940(0):157–68. [PubMed: 11458674]

Pe'er D, Hacohen N. Principles and strategies for developing network models in cancer. Cell. 2011a; 144(6):864–73. [PubMed: 21414479]

Poulikakos PI, et al. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature. 2010; 464(7287):427–30. [PubMed: 20179705]

R Core Team. R: A Language and Environment for Statistical Computing. 2013. Available at: http://www.r-project.org/

Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135(2):216–26. [PubMed: 18957198]

Ressom, H., et al. Fuzzy logic-based gene regulatory network; The 12th IEEE International Conference on Fuzzy Systems; 2003a. 2003. p. 1210-1215.FUZZ '03

Ressom H, Reynolds R, Varghese RS. Increasing the efficiency of fuzzy logic-based gene expression data analysis. Physiological genomics. 2003; 13(2):107–17. [PubMed: 12595578]

Saez-Rodriguez J, et al. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Molecular systems biology. 2009; 5(331): 331. [PubMed: 19953085]

Schulz DJ, Goaillard J-M, Marder E. Variable channel expression in identified single and electrically coupled neurons in different animals. Nature neuroscience. 2006; 9(3):356–62.

Seng TL, Bin Khalid M, Yusof R. Tuning of a neuro-fuzzy controller by genetic algorithm. IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society. 1999; 29(2):226–36.

Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003; 13(11):2498–504. [PubMed: 14597658]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sokhansanj, B. a, et al. Linear fuzzy gene network models obtained from microarray data by exhaustive search. BMC bioinformatics. 2004; 5:108. [PubMed: 15304201]

Sumners C, Fleegal MA, Zhu M. Experimental Biology 2001 Symposium Neurotransmitters in Cardiovascular Regulation: Angiotensin ANGIOTENSIN AT 1 RECEPTOR SIGNALLING PATHWAYS IN NEURONS. 2002:483–490.

Tay S, et al. Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. Nature. 2010; 466(7303):267–71. [PubMed: 20581820]

Toledo-Rodriguez M, et al. Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. Cerebral cortex (New York, N.Y.: 1991). 2004; 14(12):1310–27.

Tong RM. A control engineering review of fuzzy systems. Automatica. 1977; 13(6):559–569.

Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 2010; 28(5):511–5.

Vadigepalli R, et al. PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. Omics: a journal of integrative biology. 2003; 7(3):235–52. [PubMed: 14583114]

Veerasingham SJ, Raizada MK. Brain renin-angiotensin system dysfunction in hypertension: recent advances and perspectives. British journal of pharmacology. 2003; 139(2):191–202. [PubMed: 12770924]

Venables, WN.; Ripley, BD. Modern Applied Statistics with S. Fourth. Springer; New York: 2002. Available at: http://www.stats.ox.ac.uk/pub/MASS4

Violin JD, et al. G protein-coupled receptor kinase and beta-arrestin-mediated desensitization of the angiotensin II type 1A receptor elucidated by diacylglycerol dynamics. The Journal of biological chemistry. 2006; 281(47):36411–9. [PubMed: 17008309]

Wang J, et al. Deficiency of Microvascular Thrombomodulin and Up-Regulation of Protease-Activated Receptor-1 in Irradiated Rat Intestine. The American Journal of Pathology. 2002; 160(6):2063–2072. [PubMed: 12057911]

Woolf PJ, et al. A fuzzy logic approach to analyzing gene expression data A fuzzy logic approach to analyzing gene expression data. Physiol. Genomics. 2000; 3:9–15. [PubMed: 11015595]

Wright JW, Harding JW. The brain renin-angiotensin system: a diversity of functions and implications for CNS diseases. Pflügers Archiv: European journal of physiology. 2013; 465(1):133–51. [PubMed: 22535332]

Ying H. A general technique for deriving analytical structure of fuzzy controllers using arbitrary trapezoidal input fuzzy sets and Zadeh AND operator. Automatica. 2003; 39(7):1171–1184.

Zadeh LA. Fuzzy sets. Information and Control. 1965; 8(3):338–353.

Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology. 2005; 4(1) Article17.

Zhang S, et al. Discovering functions and revealing mechanisms at molecular level from biological networks. Proteomics. 2007; 7(16):2856–69. [PubMed: 17703505]

Zhong H, Neubig RR. Regulator of G protein signaling proteins: novel multifunctional drug targets. The Journal of pharmacology and experimental therapeutics. 2001; 297(3):837–45. [PubMed: 11356902]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

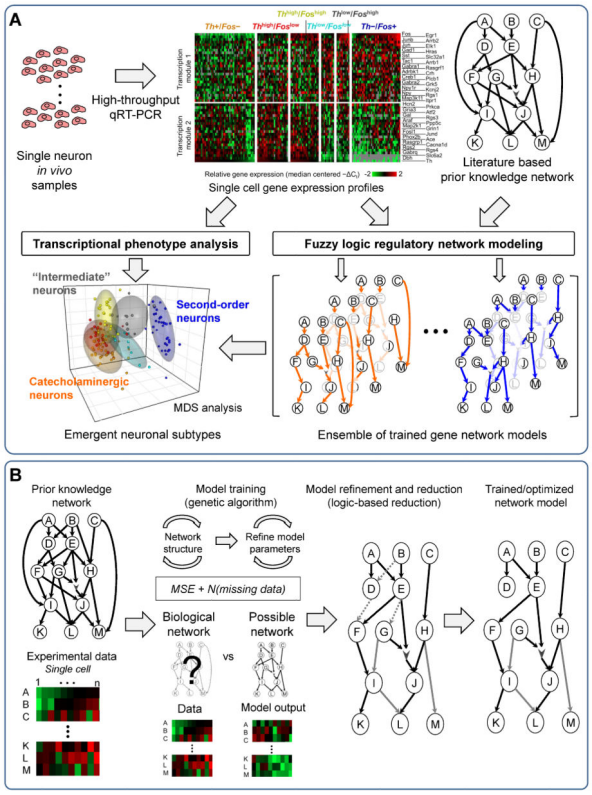

Figure 1. Inferring gene regulatory networks via fuzzy logic

(A) Transcriptional profiles of 300 individual brainstem neurons taken from hypertensive rats were measured using a high-throughput qRT-PCR platform (BioMark^{TM}). Multidimensional analysis of the single cell transcriptional data set revealed emergent neuronal subtypes that aligned with synaptic input-type and were composed of two sets of correlated gene expression modules. To infer the gene regulatory networks underlying these neuronal subtypes, an *a priori* network is trained against the single-cell transcriptional data set using principles of fuzzy logic modeling. Multiple quantitative regulatory network models are generated from the fuzzy logic methodology and are considered as part of a network ensemble that quantitatively characterizes plausible gene interactions and influences driving the highly variable transcriptional state of individual neurons. **(B)** The fuzzy logic methodology involves training an *a priori* network composed of gene interactions and influences curated from literature and transcription factor databases against context-specific gene expression data. Using a genetic algorithm, the initial *a priori* network is optimized such that unnecessary interactions (or edges) are removed while maintaining the network model's ability to fit the experimental data (below a certain error threshold). The trained network is then refined, where network model parameters (ref figure 2) are optimized to improve model fit. A final reduction step is then performed where redundant directed edges not necessary for the model to fit the experimental data are removed to generate a simplified regulatory network model.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

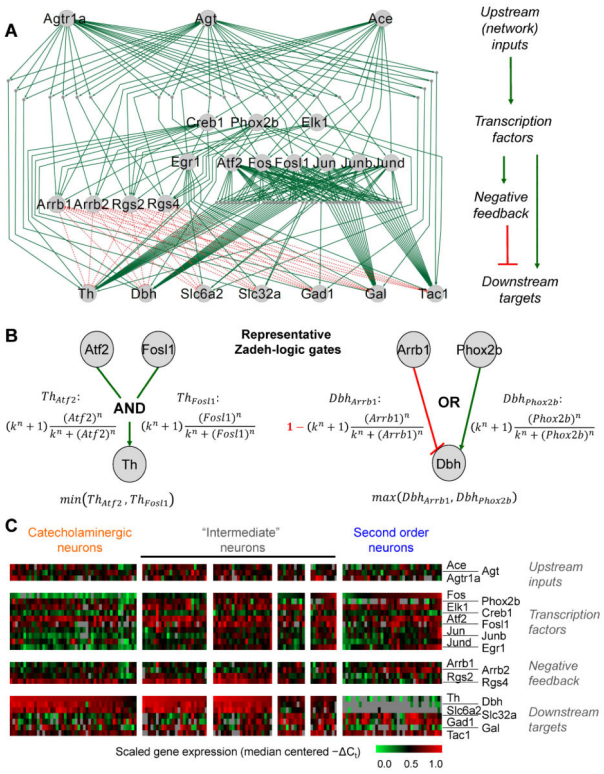

Figure 2. The *a priori* network model of the AT1R-mediated regulatory network

(A) Gene interactions and influences associated with the AT1R-mediated pathway were used to formulate the *a priori* network. The network consists of four tiers – i) upstream network inputs, ii) transcription factors, iii) negative feedback, and iv) downstream targets – which are hierarchically layered to capture the signaling cascade that occurs when AT1R is stimulated by AngII binding. Genes of interest are assigned to each tier according to their functional role in this pathway (ref Table 1). Because the transcriptional dataset consists of a single time point, the *a priori* network is formulated as a directed acyclic network where negative feedback influences, which regulate Agtr1a, are modeled as direct inhibitory interactions with downstream target genes, which are ultimately down-regulated by the influence of these negative feedback genes. **(B)** Examples of transfer functions and logic gates used in the fuzzy logic formulation of the *a priori* network. A gene interaction is modeled using a normalized Hill function, where 'n' is the hill coefficient and 'k' is the sensitivity parameter that is used to determine an output's sensitivity to an input. When a gene node has more than one input, the reaction is modeled using either an AND or OR gate. In the depicted AND gate example, the minimum value of *Th*, as calculated by the two transfer functions (Th_{Atf2} and Th_{Fosl1}), determines the resulting *Th* value. The minimum value is used in order to represent the effects a limiting substrate would have in a multi-input reaction. In the depicted OR gate example, the maximum value of *Dbh*, as calculated by the two transfer functions (Dbh_{Arrb1} and Dbh_{Phox2b}) is used to determine the resulting *Dbh* value. In this case, the maximum value is used to simulate a competitive reaction, where the input with a greater influence on the output will determine the reaction output. **(C)** Heat map of scaled $-\Delta C_t$ gene expression values used to train the *a priori* networks. Both neuronal subtype-specific data subsets and the entire $-\Delta C_t$ set were used to train the *a priori* network

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to determine if a subtype-specific or a common regulatory network would better fit and predict the experimental data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

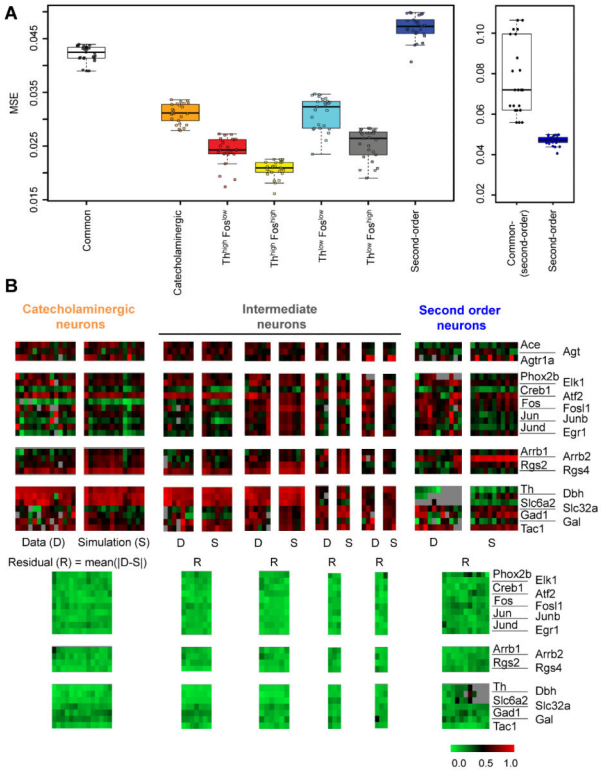

Figure 3. Trained regulatory network performance

(A) Boxplot showing MSE values, representing a network's overall ability to fit experimental data, of the general network and all neuronal subtype specific network ensembles. With the exception of the second order network ensemble, the subtype-specific network ensembles produced lower MSE values than the general network ensemble. When assessing the general network ensemble's ability to fit the second order subtype-specific data subset, however, the general network ensemble produced a higher MSE value (second boxplot). Therefore, all subtype-specific ensembles were able to fit the experimental data better than the general ensemble. **(B)** Heat maps were used to visually assess the predictive capabilities of the subtype-specific trained networks. Each pair of heat maps consists of the "validation" data subset, set-aside during the cross-validation procedure, and the corresponding simulated scaled gene expression data of single neurons from a representative trained network from the subtype-specific network ensemble. Mean absolute residuals. The absolute difference between the "validation" subset and corresponding simulated scaled gene expression of single neurons was calculated across all networks within a subtype-specific ensemble and averaged. Gray pixels represent instances where no data (failed PCR reaction) was available.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

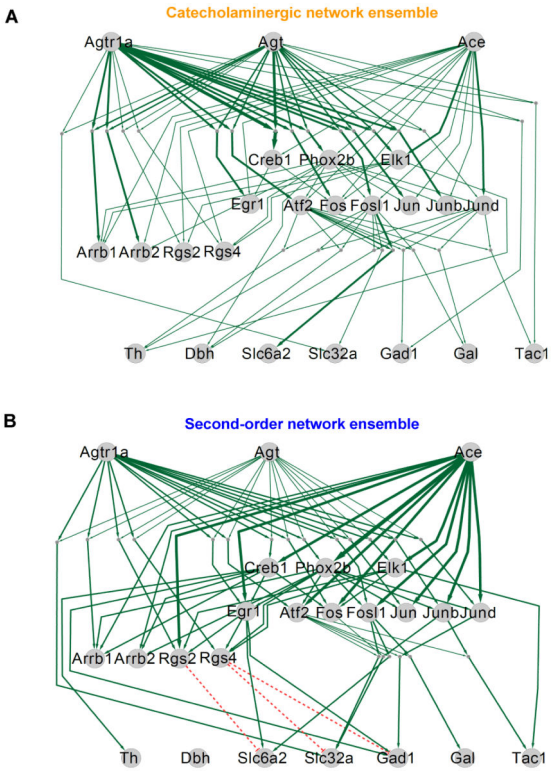

Figure 4. Comparison of higher-order and second order neuronal network ensembles

The unprocessed-refined-reduced network ensembles underlying the **(A)** catecholaminergic and **(B)** second-order neuronal subtypes are depicted in the first and second panel. Edge thickness represents the strength (i.e. frequency) of a particular interaction as determined during the training process and across all networks within an ensemble. Edge color represents either stimulatory (green) or inhibitory (red) interactions. A significant amount of interactions are pruned from the original *a priori* network structure, particularly in the interactions between transcription factors and downstream target genes. Edge target shape defines the interaction type–stimulatory (arrowhead) or inhibitory (T-head). Common edges are depicted as grey colored edges. The frequency of edges (x) was classified into three bins: i) $0.12 \leq \times \leq 0.4$, ii) $0.4 < \times \leq 0.7$, and iii) $0.7 < \times \leq 1.0$. Only edge frequencies ≥ 0.12 were included in network ensemble visualization in order to focus on more dominant interactions and aid in visual interpretation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

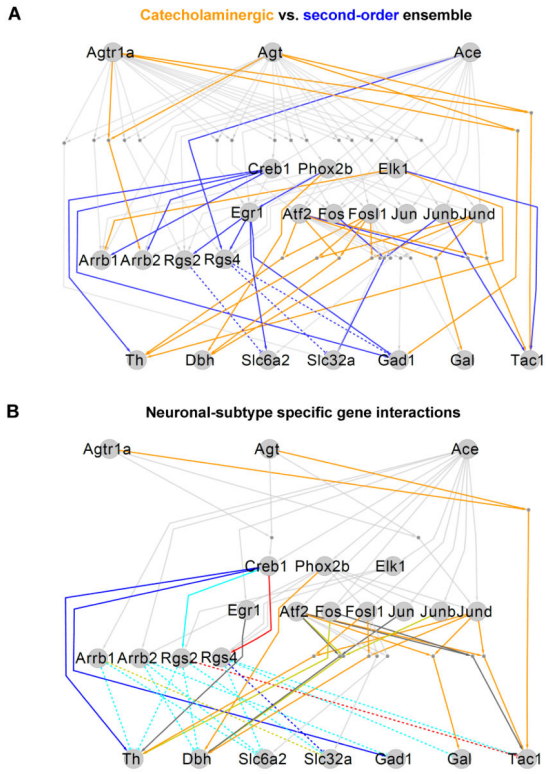

Figure 5. Distinct gene regulatory networks distinguish neuronal subtypes

(A) The network depicts both common and unique edges between the catecholaminergic and second-order network ensembles. Gene interactions unique to catecholaminergic neurons (orange) depict many more AP-1 transcription factor related interactions on *Th* and *Dbh* while interactions specific to second-order neurons (blue) depict distinct transcription factor regulation of downstream target genes. **(B)** As neurons traverse the intermediate neuronal subtypes that lie along the transcriptional landscape (Park, Brureau et al. 2014), distinct causal gene interactions are driving these transcriptional states. Unique directed edges for the intermediate neuronal subtypes are colored as follows: Th^{high} Fos^{low} (red), Th^{high} Fos^{high} (yellow), Th^{low} Fos^{low} (cyan), Th^{low} Fos^{high} (dark grey). Directed edges common to all neuronal subtypes are shown in light grey. Edge frequency is not accounted for in the network visualizations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

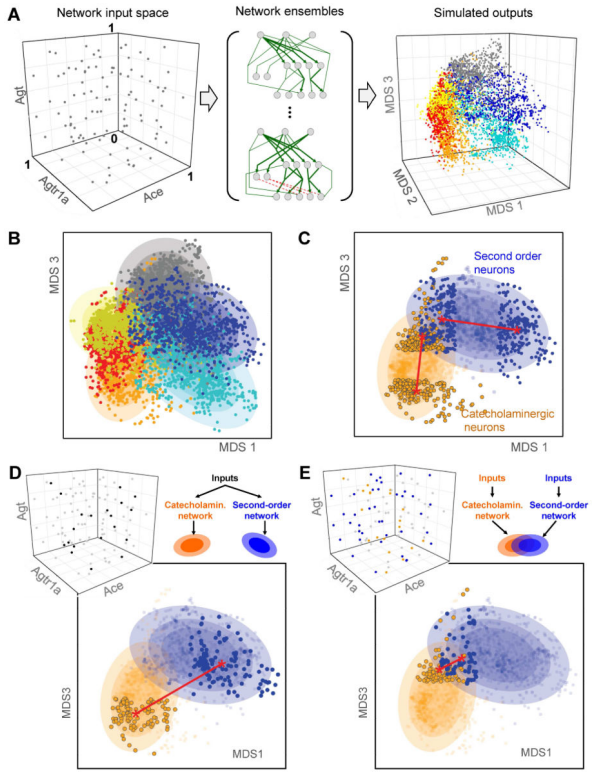

Figure 6. Interpreting single cell variability via multidimensional scaling of network simulations

(A) Workflow of network ensemble simulations and analysis via multidimensional scaling (MDS). Random network input values (scaled gene expression values) for *Ace*, *Agt*, and *Agtr1a* were selected from the multidimensional network input space using Latin hypercube sampling. This set of randomly selected network input values was used to stimulate all subtype-specific network ensembles. Using our previous analytical approach (Park, Brureau et al. 2014), the resulting transcriptional state for simulated single cell was then projected into a 3D space using MDS. Briefly, all pairwise dissimilarity values [−1 1] were calculated for all simulated neurons, based on their multidimensional transcriptional profile, using the Spearman rank correlation. The dissimilarity matrix is then projected into a lower dimensional space using a non-metric MDS technique. The projected position of each point in the lower-dimensional space is iteratively determined until a minimum difference between the pairwise-distance in the lower-dimensional space and original-dimensional space is minimized. Thus, the greater the distance between any two points in the MDS space, the greater the difference is between their ranked transcriptional states. Similar to the emergent and previously identified neuronal subtypes, distinct neuronal subtypes emerge from network ensemble simulations. The colored contour regions represent population density for each neuronal subtype where the inner, middle, and outer ellipses represent the 65%-, 80%-, and 95%- quantile respectively. **(B)** The first and second MDS coordinates capture the variability across single neurons and the third MDS coordinate captures variability due to the emergent neuronal subtypes. Subsequent analysis was performed along MDS coordinate axes 1 and 3. **(C)** Significant variability is reproduced in the network ensemble simulations, as indicated by the spread of the data points within and across neuronal subtypes. An overlap between neuronal subtypes, however, is also observed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Focusing on the higher-order and second-order neuronal subtypes, the intra-subtype Euclidean distance between simulated neurons within the upper 75%- and lower 25%-quantile of the population density (along MDS axis 3) is much larger than the distance between higher-order neurons in the 75%-quantile and second-order neurons within the lower 25%-quantile of the population density (along MDS1) (0.3391 vs. 0.1764). This larger intra-subtype distance also holds for the second-order neuronal subtype (0.5481 vs. 0.1764). Distances are determined between centroids for each quantile-group. To explore what is contributing to this variability, transcriptional states generated from distinct network inputs stimulating the catecholaminergic (catecholamine.) and second-order regulatory network ensembles were examined. **(D)** Subtype-specific network responses to similar network inputs yielding distinct transcriptional states. **(E)**. Second-order subtype-specific network response to distinct network inputs yield similar transcriptional states.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

AT1R a *priori* regulatory network genes

Upstream (network) inputs • *Ace, Agt, Agtr1a* – (Sumners et al. 2002; Mehta & Griendling 2007)
Transcription factors • *Phox2b, Egr1, Atf2* – (Kvetnansky et al. 2009) • *Fos, Jun* – (Kvetnansky et al. 2009; Sumners et al. 2002) • *Fosl1, Junb, Jund* – (Kvetnansky et al. 2009; Herdegen & Leah 1998) • *Creb1* – (Kvetnansky et al. 2009; Karin et al. 1997; Herdegen & Leah 1998) • *Elk1* – (Kvetnansky et al. 2009; Sumners et al. 2002; Karin et al. 1997)
Negative feedback • *Arrb1, Arrb2* – (Luttrell & Lefkowitz 2002; Violin et al. 2006) • *Rgs2, Rgs4* – (Zhong & Neubig 2001)
Downstream targets • *Th, Dbh* – (Veerasingham & Raizada 2003; Sumners et al. 2002) • *Slc32a* – (Gasnier 2004; Vadigepalli et al. 2003) • *Slc6a2* – (Kvetnansky et al. 2009; Sumners et al. 2002; Veerasingham & Raizada 2003) • *Gad1* – (Paton et al. 2001) • *Gal* – (Kvetnansky et al. 2009; Sumners et al. 2002; Melander et al. 1986) • *Tac1* – (Marc & Llorens-Cortes 2011; Hall et al. 1989; Wright & Harding 2013)