Inputs drive cell phenotype variability

James Park
*Thomas Jefferson University*

Anthony Brureau
*Thomas Jefferson University*

Kate Kernan
*Washington University School of Medicine in St. Louis*

Alexandria Starks
*Thomas Jefferson University*

Sonali Gulati
*Thomas Jefferson University*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

Please let us know how this document benefits you.

**Recommended Citation**

Park, James; Brureau, Anthony; Kernan, Kate; Starks, Alexandria; Gulati, Sonali; Ogunnaike, Babatunde; Schwaber, James; and Vadigepalli, Rajanikanth, "Inputs drive cell phenotype variability." *Genome Research*. 24, 930-941. (2014).

[https://digitalcommons.wustl.edu/open_access_pubs/2989](https://digitalcommons.wustl.edu/open_access_pubs/2989)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Authors
James Park, Anthony Brureau, Kate Kernan, Alexandria Starks, Sonali Gulati, Babatunde Ogunnaike, James Schwaber, and Rajanikanth Vadigepalli
Inputs drive cell phenotype variability

James Park, Anthony Brureau, Kate Kernan, et al.

Genome Res. 2014 24: 930-941 originally published online March 26, 2014
Access the most recent version at doi:10.1101/gr.161802.113

Supplemental Material
http://genome.cshlp.org/content/suppl/2014/04/09/gr.161802.113.DC1.html

References
This article cites 62 articles, 18 of which can be accessed free at:
http://genome.cshlp.org/content/24/6/930.full.html#ref-list-1

Creative Commons License
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Genome Research go to:
http://genome.cshlp.org/subscriptions

© 2014 Park et al.; Published by Cold Spring Harbor Laboratory Press
Inputs drive cell phenotype variability

James Park,1,2,4 Anthony Brureau,1,4 Kate Kernan,3 Alexandria Starks,1 Sonali Gulati,1 Babatunde Ogunnaike,2 James Schwaber,1,2,5 and Rajanikanth Vadigepalli1,2,5

1Daniel Baugh Institute for Functional Genomics and Computational Biology, Department of Pathology, Anatomy and Cell Biology, Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA; 2Department of Chemical and Biochemical Engineering, University of Delaware, Newark, Delaware 19716, USA; 3Department of Pediatrics, Washington Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA; 4Department of Chemical and Biochemical Engineering, University of Delaware, Newark, Delaware 19716, USA; 5Department of Pediatrics, Washington University Saint Louis, Saint Louis, Missouri 63130, USA

What is the significance of the extensive variability observed in individual members of a single-cell phenotype? This question is particularly relevant to the highly differentiated organization of the brain. In this study, for the first time, we analyze the in vivo variability within a neuronal phenotype in terms of input type. We developed a large-scale gene-expression data set of data from several hundred single brainstem neurons selected on the basis of their specific synaptic input types. The results show a surprising organizational structure in which neuronal variability aligned with input type along a continuum of sub-phenotypes and corresponding gene regulatory modules. Correlations between these regulatory modules and specific cellular states were stratified by synaptic input type. Moreover, we found that the phenotype gradient and correlated regulatory modules were maintained across subjects. As these specific cellular states are a function of the inputs received, the stability of these states represents “attractor”-like states along a dynamic landscape that is influenced and shaped by inputs, enabling distinct state-dependent functional responses. We interpret the phenotype gradient as arising from analog tuning of underlying regulatory networks driven by distinct inputs to individual cells. Our results change the way we understand how a phenotypic population supports robust biological function by integrating the environmental experience of individual cells. Our results provide an explanation of the functional significance of the pervasive variability observed within a cell type and are broadly applicable to understanding the relationship between cellular input history and cell phenotype within all tissues.

[Supplemental material is available for this article.]

What is a cell type? This question has been a central project of biology and molecular biology. Typically, we deconstruct a tissue or organ into its constituent cell types based on anatomical, physiological or biochemical features, and examine each distinct cell phenotype to understand its larger function. In this context, it has been a major biological aspiration to connect cell phenotype to the genome via gene expression. But elucidating the organization of cell types by linking cell phenotype analysis to transcriptional state has been largely elusive. This elusiveness is due to the variability seen in transcriptional data sets produced from what are expected to be homogeneous cell populations. As high-throughput data acquisition methods have now become highly precise, it has become obvious that the variability observed in the results is not a mere distribution around a mean, but reflects true heterogeneity, the activity of cells in a range of distinct states. Even when we take single-cell genomic measures, this variability within cell types persists (Guo et al. 2010; Eberwine and Bartfai 2011). This variability is not only present at the individual cell level, but extends to the levels of electrical and neural network function (Eberwine and Bartfai 2011; Marder 2011; Marder and Taylor 2011). Reconciling cell type in the face of such heterogeneity in the adult mammalian brain and accurately defining post-development diversity continue to be difficult challenges, as expressed by Birren and Marder (2013) and Wichterle et al. (2013). The significance of understanding and defining brain cell type is highlighted in the recently announced BRAIN Initiative, which seeks to understand how “cellular phenotypes based on transcriptional profiles may change as a function of developmental stage, age, cell state (e.g., cell cycle for mitotic cells), activity levels, and experience among other things” (NIH RFA-MH-14-215 2013). We believe our results are part of a solution to this problem.

Defining cell type within the highly differentiated and networked mammalian brain relies on location, connectivity, morphology, histochemistry, neurotransmitter type, and most recently on transcriptomic profiles. Significant efforts have detailed how coordinated transcriptional mechanisms lead to neuronal diversification and connectivity in the context of developmental dynamics (Chen et al. 2006b; Kramer et al. 2006; Luo et al. 2008; Friese et al. 2009). However, an increasing amount of evidence demonstrates significant heterogeneity and plasticity caused by further post-developmental, adaptive changes within developed lineages. Cells remain plastic and are able to change adaptively in response to inputs; rather than reaching a final stable state or cell fate they continue to acquire new response capabilities in the mature organism. Thus, the current state of a cell is a product of the cumulative influences or inputs received throughout its history. Recent results support the idea that this cumulative record is represented by the transcriptome, representing an essential “snapshot state memory” of the phenotype (e.g., Kim and Eberwine 2010).

*These authors contributed equally to this work.

*Corresponding authors
E-mail james.schwaber@jefferson.edu
E-mail rajanikanth.vadigepalli@jefferson.edu

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.161802.113.
The cell’s transcriptome adapts to inputs to change the cell, in effect becoming a repository of the cell’s input history.

In the context of mature neurons, recent experiments demonstrate how cellular experience influences heterogeneity through “neurotransmitter respecification” in adult rat brains, which was accomplished by modifying the amount of light/dark stimulus received by these adult rats (Dulcis et al. 2013). Another example shows in vivo reprogramming of circuit connectivity in mature neocortical neurons in mice (De la Rossa et al. 2013). As cutting edge discovery of plasticity and diversity within and across neuron types continues, the causes of these phenomena remain unclear (De la Rossa et al. 2013; Dulcis et al. 2013). We suspect that adaptive responses to inputs of this kind may cause the variability that is observed in high-throughput studies of phenotypically similar cells (Eberwine and Bartfai 2011; Kim et al. 2011). In other words, a cell type that might have been expected to be homogenous, sharing a common end fate, might rather be heterogeneous due to each cell within the cell type adapting to a distinct input history. Therefore, we hypothesize that neuronal transcriptomic variability reflects synaptic input variability to the phenotype’s individual cell members. We test this hypothesis by examining gene expression differences within individual mature neurons of the same neuroanatomical phenotype. We analyze the potential organization of these differences in terms of neuronal input types. If such an organization were supported by the data, expression variability would be functionally meaningful by facilitating alternative responses within the phenotype.

We now have been able to investigate this, for the first time, by taking in vivo measures of each cell’s high-throughput transcriptional state in several hundred neurons taken from a single nucleus. Specifically, we investigated this hypothesis by studying the nucleus tractus solitarius (NTS), a brainstem nucleus consisting of the anatomical phenotype of medium sized, fusiform neurons forming a column in the dorsal medulla as the sensory nucleus for the viscera. The neurons surround and are innervated by the tractus solitarius (ts) that conveys afferent inputs of the glossopharyngeal and vagus cranial nerves from visceral organs, including blood pressure sensory baroreceptor afferents. The NTS plays an integrative role in autonomic homeostasis, receiving inputs that place demands on homeostasis such as inputs conveying visceral states, pain, posture, exercise, temperature, circadian time, and mood. As individual NTS neurons must integrate distinct combinatorial input sets, their variability may reflect their inputs. We examined the NTS phenotype in vivo in its native tissue context using microfluidic qPCR across several hundred NTS neurons (Supplemental Material and Methods). In these single neurons we assayed the expression of 96 key genes derived from a previous microarray study of the nucleus (Khan et al. 2008).

NTS neurons were identified by two cell markers signifying distinct inputs, FOS and tyrosine hydroxylase (TH) (Fig. 1). A very extensive amount of literature demonstrates the use of immediate early gene Fos as an indicator for the subset of NTS neurons responsive to acute hypertensive disturbances. FOS is an established indicator of cellular activation—in the present case identifying cells directly influenced by blood pressure baroreceptor afferent inputs (Li and Dampney 1992, 1994; Rogers et al. 1993; Chan and Sawchenko 1994, 1995, 1998; Glass et al. 2007). Carefully conducted control studies have shown that high Fos levels depend on neurons being directly influenced by the increased activity of the blood pressure baroreceptor afferent inputs (Potts et al. 1997; Chan et al. 2000). Simultaneously, the NTS population of norepinephrine cells, indicated by the expression of the catecholamine synthesis enzyme TH, receives “higher order” influences through one or more additional interneurons and integrative inputs rather than direct blood pressure inputs from baroreceptors afferents (Chen and Sawchenko 1994, 1995, 1998; Rinaman 2010, 2011). Thus, the markers for FOS and TH identify discrete NTS neuronal populations in term of their expected inputs. Using Fos and Th as markers for distinguishing the expected input types, we investigated the differences in transcriptional states of individual NTS neurons with respect to these inputs.

Results

We obtained a high-dimensional single neuron gene expression data set comprised of 28,880 data points representing expression of 96 genes each in 300 single neurons lifted from the NTS of six rats (Supplemental Material and Methods). Gene expression levels were measured using a high-throughput qPCR platform (BioMark, Fluidigm), which has demonstrated the ability to reproducibly measure gene expression over five orders of magnitude with minimal technical variability (Supplemental Fig. S1). Our quality control workflow followed established procedures for minimizing nonspecific contamination of samples (Espina et al. 2006). In addition to visual inspection of tissue and captured cell body, we measured neuron, astrocyte, microglial, and endothelial specific gene expression in a separate set of single neurons and astrocytes collected from the NTS via laser capture microdissection (LCM) to test for potential cross-contamination from nontargeted cell types that may affect our single cell samples (Supplemental Material and Methods). Our results showed minimal to nonexistent crossover contamination (Fig. 2; Supplemental Figs. S2, S3) consistent with the repeated performance of LCM approaches used by a number of other groups (e.g., Wang et al. 2002; Ye et al. 2003; Zhang et al. 2003; Espina et al. 2006; Macdonald et al. 2008). We normalized the high-dimensional single neuron gene expression data using established approaches that evaluated multiple reference genes to select those suitable for data normalization and comparison across single cells (Supplemental Fig. S4; Vandesompele et al. 2002; Andersen et al. 2004).

Our results revealed significant variability in normalized gene expression across all single cells (Fig. 3). Approximately two-thirds of the genes showed expression values spanning three orders of magnitude as measured over multiple high-throughput qPCR runs, multiple rats, and in hypertensive and baseline states. Initially we analyzed the variability in single cell gene expression using Principal Component Analysis (PCA). Our results revealed an unstructured scatter of cellular states, as shown by the PCA scores along the first five principal components that accounted for 48.94% of the variability in the data (Fig. 1B,C). We derived a subset of 48 genes that significantly contributed to the observed variability, using the five highest and lowest corresponding loading values along the first five principal components as a basis for gene selection (Fig. 1B; Supplemental Fig. S5; Supplemental Table S4; Supplemental Material and Methods). Next, we analyzed the single cell states, characterized by this 48-dimensional gene expression (i.e., transcriptional) profile per cell, to determine the presence of any structure or organization subtending these seemingly disparate cell states. We performed all possible pairwise comparisons of the single cells using a Spearman rank correlation. The high-dimensional data set of correlation coefficients (Fig. 1D) were then converted to corresponding similarity values, which were projected into three dimensions using multidimensional scaling (MDS) (Fig. 1E). MDS is a technique useful to visualize...
ceptor afferent inputs, and Fos surrogates for different neuronal response capacity to particular variability observed in the data. (highest loading values along PC 2 are labeled. The highest and lowest genes along the multiple PCs components were explored (Supplemental Information). (of single cells from hypertensive rats was performed to analyze variation within gene expression data. We subsequently analyzed the single cell variability with respect to gene expression of the two input-type markers Fos expression denotes catecholaminergic neurons receiving “higher order” integrative inputs. We first considered the extremes of the single cell multiplex gene expression distribution (Fig. 1E) with respect to the two input-type markers to identify and annotate two input-based subtypes (Fig. 4A); cells with Fos expression and minimal to no Th expression (Th−/Fos+) and cells with Th expression and minimal to no Fos expression (Th+/Fos−). Categorization of the same single cells based on mRNA expression or immunoreactivity to the respective markers was nearly identical. Only six single cells categorized as “Fos−” showed FOS immunoreactivity. Although a single cell may be labeled as “Fos−”, this annotation is simply an indicator of low Fos mRNA levels present in that particular cell. Given the dynamic and transient nature of Fos regulation, this slight discrepancy is unsurprising.

Applying the mRNA-based annotation of these single cell subtypes to the MDS visualization of cell states revealed a surprisingly structured organization. The two subtypes were distinctly clustered at the opposing extremes of the overall distribution of cells (Fig. 4B; Supplemental Fig. S6A,C,E). The separation of the two extreme subtypes was statistically significant as no such clustering was observed in randomized permutations of the data (Supplemental Fig. S7). Cells categorized by their input types (i.e., Fos or Th expression level) maintained close proximity to each other in the transcriptional space, indicating that individual cells receiving a particular input type share similar transcriptional profiles, an indicator of cell response.

This structured organization supports a novel perspective that differential inputs to individual cells may drive variation in the transcriptional profiles of NTS neurons. It is interesting to note that the highly variable genes identified using PCA were rank-correlated across many single cells categorized by the two input-type markers (Fig. 4C). The underlying gene expression was organized into two correlative modules with the Fos and Th expression profiles serving as exemplars for each group to distinctly separate the two populations of NTS cells (Fig. 4C,D). Note that the expression of other key genes relevant to catecholaminergic function (e.g., Dbh and Slc6a2) was most highly correlated with Th gene expression, consistent with a well-regarded expectation of co-regulation of these genes (Qadri et al. 1991; Stadler et al. 1992; Lu et al. 1996; Blume et al. 1999; Richards et al. 1999; Gallinat 2001; Dogan et al. 2004). This result serves as an internal validation of our analysis. The alignment of the two input-type markers with the variation seen in the measured transcriptional profiles of NTS cells implies a causal relationship where inputs to individual cells play a major role in similarities and dissimilarities of high-dimensional data in a lower-dimensional space (Fuller et al. 2002; Ross et al. 2003; Taguchi and Oono 2005). In this context, the proximity between any two cells in the MDS space corresponds to how similar or dissimilar the rank correlation of gene expression is between that particular cell pair. The single cells were distributed in a cloud in this MDS mapping (Fig. 1E), revealing no initially obvious structures or organization to cell states.

We subsequently analyzed the single cell variability with respect to gene expression of the two input-type markers Fos (transcript of Fos) and Th. Both Fos and Th expression levels are surrogates for different neuronal response capacity to particular inputs; Fos expression denotes neurons directly receiving baroreceptor afferent inputs, and Th expression denotes catecholaminergic neurons receiving “higher order” integrative inputs. We first considered the extremes of the single cell multiplex gene expression distribution (Fig. 1E) with respect to the two input-type markers to identify and annotate two input-based subtypes (Fig. 4A); cells with Fos expression and minimal to no Th expression (Th−/Fos+) and cells with Th expression and minimal to no Fos expression (Th+/Fos−). Categorization of the same single cells based on mRNA expression or immunoreactivity to the respective markers was nearly identical. Only six single cells categorized as “Fos−” showed FOS immunoreactivity. Although a single cell may be labeled as “Fos−”, this annotation is simply an indicator of low Fos mRNA levels present in that particular cell. Given the dynamic and transient nature of Fos regulation, this slight discrepancy is unsurprising.

Applying the mRNA-based annotation of these single cell subtypes to the MDS visualization of cell states revealed a surprisingly structured organization. The two subtypes were distinctly clustered at the opposing extremes of the overall distribution of cells (Fig. 4B; Supplemental Fig. S6A,C,E). The separation of the two extreme subtypes was statistically significant as no such clustering was observed in randomized permutations of the data (Supplemental Fig. S7). Cells categorized by their input types (i.e., Fos or Th expression level) maintained close proximity to each other in the transcriptional space, indicating that individual cells receiving a particular input type share similar transcriptional profiles, an indicator of cell response.

This structured organization supports a novel perspective that differential inputs to individual cells may drive variation in the transcriptional profiles of NTS neurons. It is interesting to note that the highly variable genes identified using PCA were rank-correlated across many single cells categorized by the two input-type markers (Fig. 4C). The underlying gene expression was organized into two correlative modules with the Fos and Th expression profiles serving as exemplars for each group to distinctly separate the two populations of NTS cells (Fig. 4C,D). Note that the expression of other key genes relevant to catecholaminergic function (e.g., Dbh and Slc6a2) was most highly correlated with Th gene expression, consistent with a well-regarded expectation of co-regulation of these genes (Qadri et al. 1991; Stadler et al. 1992; Lu et al. 1996; Blume et al. 1999; Richards et al. 1999; Gallinat 2001; Dogan et al. 2004). This result serves as an internal validation of our analysis. The alignment of the two input-type markers with the variation seen in the measured transcriptional profiles of NTS cells implies a causal relationship where inputs to individual cells play a major role in
shaping the transcriptional profiles. This relationship argues that inputs influence neuronal transcriptional states and is further substantiated by the quantitative nature of this relationship. Regardless of the defined threshold for Fos expression or Th expression, cells with the highest Fos or highest Th expression tended to be the same cells having extreme expression of the 48 variable genes highlighted by PCA (Supplemental Figs. S8–S11).

While the majority of the cells expressed either Fos or Th, there were smaller populations with lower expression of one or both input-type markers. We interpret these various expression levels as indications of different populations with respect to the hypertensive challenge, the relationships between TFs and genes within the transcription module were reduced and shift mainly to genes in transcription module 2 in the Th+/Fos− (higher-order input cell) and Th−/Fos+ (second-order input cell) networks (Fig. 6).

These input-driven shifts in expression correlation and potential gene regulation effects were also reflected in the constrained space occupied by hypertensive Th+/Fos− cells relative to the Th+ cells from control animals at baseline blood pressure levels (Supplemental Fig. S13). Although there was some individual rat-to-rat variability within these transcriptional modules, the same two input types. For example, some subset of Th-expressing cells may respond weakly to baroreceptor inputs through interactions with interneurons, yielding variable Fos expression in those cells. If so, the NTS cell types may form a continuous distribution with respect to strength of input from different sources, and by implication a continuous distribution of expression patterns may result. With this expectation, we categorized cells with lower expression of Th and Fos based on median expression for each input-type marker to yield four “intermediate” subtypes (Fig. 5A; Supplemental Material and Methods, section “Single Cell Subtypes”). Mapping these annotations onto the MDS visualization of cell states revealed that the subtypes showing lower Th or Fos levels were located in between the two extreme cell types (Fig. 5B). Similarly, subtypes showing higher levels of Th or Fos aligned closer to the corresponding extreme input-based subtype. Additionally, these results indicate that the 48 highly variable genes show correlated expression within these intermediate subtypes (Fig. 5C). The gene expression in the intermediate subtypes was correlated based on the same modules observed in the cases of the extreme Th+/Fos− and Th−/Fos+ subtypes (Supplemental Fig. S12).

The coordinated gene expression patterns, or transcription modules (Fig. 4C), further differentiate the expression states of cells. The various active states that lie along the gradient structure are governed by underlying gene regulatory networks, which can be used to further distinguish these states. A comparison of rank correlative gene networks, a surrogate for the regulatory interactions occurring in single cells within the extreme states of Figure 4, shows distinct structures of correlative gene expression behavior (Fig. 6). In the baseline Th+/Fos− network, transcription factor (TFs) showed a high degree of connectivity, i.e., correlative relationships, with genes across both transcription modules. However, under the

**Figure 2.** Single neuron and astrocyte LCM. (A) Tyrosine hydroxylase (TH) immunohistochemical staining and collection of TH+ single cells from a coronal section of a normotensive rat brainstem. Colored outline images represent magnified tissue sections from which TH+ single cells were captured. (B) Glial fibrillary acidic protein (GFAP) immunohistochemical staining and collection of GFAP+ single cells from an adjacent coronal section of a normotensive rat brainstem. Colored outline images represent magnified tissue sections from which GFAP+ single cells were captured. (C) Gel electrophoresis image of reverse-transcribed cDNA from whole-brain tissue (positive control) (lanes 1–3), a representative single neuron sample (lanes 5–7), a representative astrocyte sample (lanes 8–10), and a no-template control (NTC) (negative control) (lanes 11,12). All samples underwent 22 pre-amplification cycles prior to undergoing a 40-cycle PCR. Products from the 40-cycle PCR were placed on an E-Gel EX Agarose Gel 4% (Invitrogen). The rat whole brain positive control shows product bands for Gapdh (<50 bp suggests a nonspecific product. Similar behavior is observed in lane 9, where the astrocyte sample shows no Th product at the expected 68-bp size. Only a light product band at <50 bp is present, suggesting a nonspecific product. The results indicate minimal to no crossover contamination occurring between astrocytes and neurons.
pattern of structured variation across input classes was present in each animal (Supplemental Fig. S14).

Discussion

Having analyzed NTS neurons of the same neuroanatomical phenotype in vivo in the context of their specific input connectivity, we found that post-developmental neuronal cell type is strongly associated with the specificity of connections. Studying gene expression profiles of NTS neurons at the single cell level provided us with the appropriate resolution to distinguish cell types with respect to the inputs they received. Our results support the importance of connectivity in defining a cell type, through the transcriptional regulation of neurons by their inputs. Viewing the distribution of neuronal cell types as a function of specific inputs allowed us to interpret cell-to-cell variability as structured heterogeneity rather than noise around a mean.

This single cell variability likely reflects cellular functional heterogeneity (Enver et al. 2009), influencing a cell’s position along the gradient of the observed multiplex gene expression (Fig. 4D). This structure is evident in the MDS visualization where single cells fall into input-defined clusters of cells that are positioned along an expression pattern gradient (Figs. 4D, 5C). Since input history of an individual cell influences the cell’s transcriptomic state, we postulate that the cumulative input history of a cell provides a driving force for adjustment or analog tuning of the transcription modules, placing cells within interchangeable, stable states along the gradient of catecholaminergic (Th+/Fos-) and non-catecholaminergic (Th−/Fos+) hypertension responsive cell states.

Visualization, using MDS, of gene expression gradients, dynamic landscapes, and analog tuning of expression defining cell development and function is a recent application used most notably in hematopoietic and embryonic stem cells and cell signaling systems, such as NFkB signaling (Hough et al. 2009; Tay et al. 2010; Bendall et al. 2011). Our application of such techniques and concepts to ostensibly terminally differentiated single cells is novel as far as we are aware. The input-based ordered structure within the

Figure 3. Gene expression and variance distributions. Boxplots overlaid with in-line scatter plots showing the spread of expression data for all genes (ΔCq). Each gray dot corresponds to a particular gene expression level in a particular single cell sample. (A) Baseline-normotensive cells. (B) Hypertensive cells.
heterogeneous gene expression of single neurons in the MDS space now allows us to contextualize single cells along transcriptional module gradients, suggesting a plastic rather than a discrete cell phenotype. Finding correlated gene expression modules delineated by inputs is consistent with transcriptional phenotypes that result from combinatorial inputs. Subsequent variability within a given phenotype results from differences in input type and strength to each cell. In this context, any additional variability within a sub-phenotype, reflected in the spread of single cells of that particular group, may reflect variability of other inputs to the cell population. Additional input-driven analysis would be expected to further fractionate the phenotype.

Our results, which suggest an input-based organization of the NTS neuronal phenotype within a cloud of cellular states, raise intriguing possibilities as to the mechanisms through which such a gene expression gradient could be tuned in individual neurons. It is likely that combinatorial actions of transcriptional and post-transcriptional regulatory processes are involved in transducing cellular inputs into the downstream regulation of transcriptional states. Such regulatory network coordination to generate complex patterns of gene expression has been well described with respect to developmental dynamics, and typically involves a unique combination of regulatory factors for each cell type (Chen et al. 2006a,b; Kramer et al. 2006; Luo et al. 2008; Friesie et al. 2009). It is possible that such formalism extends into post-developmental gene expression variability between neuronal phenotypes. We should also consider alternative regulatory schemes where graded gene expression spanning the spectrum of cellular states may be driven by a set of regulators in common with the NTS neuronal phenotypes, with inputs tuning cell-to-cell differences in regulatory transitions.

Figure 4. Input-driven extreme phenotypes. (A) Input-based cell type identification. A bivariate plot of single cells obtained from hypertensive rats based on their normalized expression of Th and Fos. The extremes of the distributions for each gene were initially explored, resulting in two extreme classes of single cells: (1) cells with no Th expression and (2) cells with no fos expression. The gray lines indicate the threshold criteria used to define the extreme subtypes Th-/fos+ (blue filled circles) and Th+/fos- (orange filled circles). The effect of various thresholds used for classifying extreme classes is shown in Supplemental Figures S8–S11. (B) Clustering of input-based cell types. Three-dimensional MDS projection of single cells with the extreme phenotype classifications applied. These projects are based on the similarity of single cells with respect to their ranked expression order of the 48 gene subset. Alternative viewing perspectives of the 3D MDS projections are included in Supplemental Figure S6A,C,E. (C) Gene expression correlation modules. Gene-to-gene pairwise Spearman rank correlation coefficients across the extreme subtype cells were calculated and are represented in the heat map. The highly variable genes show that the underlying gene expression in these extreme subtypes can be organized into two combinatorial groups, or transcription modules. These transcription modules group genes that show higher correlations (upper left quadrant and lower right quadrant of the heat map) with each other across single cells of the extreme subtypes. Columns and rows with the same index representing a particular gene follow the row annotation in panel D. (D) Gene expression gradients in input-based cell types. Heat map of normalized gene expression data. An overall gene expression gradient pattern can be observed in the gene expression profile of the 48 highly variable genes across the extreme subtypes. Focusing on the extreme regions to the left and right of the whitened-out region on the heat map, opposite expression behaviors can be observed in the two transcription modules between the two extreme input-based subtypes. The upper set of genes in the heat map shows an overall decrease in gene expression in Th+/fos- cells and an increase in Th-/fos+ cells. The opposite behavior is observed in the lower set of genes. The expression patterns of these extreme subtypes occupy opposite ends of the gene expression gradient observed.
activity and combinatorial action. Given the 1-h duration of the hypertension perturbation in our study, it is unlikely for the transcriptional regulatory network to influence neuronal network connectivity in such a short period. Hence, such feedback cannot serve as an alternative explanation of association between cellular activity and combinatorial action.

These interchangeable cell states can be schematically represented as a dynamic gene expression landscape populated by individual cells based on their transcriptional response (Fig. 7A,B). The landscape figures and two-dimensional contour plots are used to help illustrate the distinct cell states and the influence of inputs. Such a conceptualization is an evolution of the Waddington “canalization” to describe developmental phenotypes (Waddington 1942) and was used to organize the interrelationships between various cell types that emerge through dynamic expression changes during development (Enver et al. 2009). The contour plots are a projection of the single cells in the 3D MDS space onto a 2D plane (Fig. 7A,B; Supplemental Material and Methods, section “Contour Plots and Dynamic Landscape”). The “depth” of a well along the landscape (Supplemental Material and Methods, section “Contour Plots and Dynamic Landscape”) at any given location was derived from the local density of cells so that a cluster of many cells is deeper and indicates a potential local “attractor” reflecting constrained gene expression in those particular cells. In this representation, these valleys and wells, or “attractor”-like states, correspond to dominant expression states of relatively stable expression modules (e.g., those corresponding to \(Th+/Fos^{-}\) and \(Th^{-}/Fos^{+}\) extreme subtypes). The remaining topography corresponds to potential intermediary states that may be transient in response to input histories of individual cells and physiological perturbations (Fig. 7A,B, color-coded groupings). The path that these cells take along the gene expression landscape is a function of the input(s) received and is likely to be as varied as the input(s) (Fig. 7C).

The exposure to a hypertensive challenge changes the constraints (Supplemental Fig. S13) and distribution of cells within the gene expression landscape (Fig. 7A,B), consistent with phenotypes that are determined by distinct state-dependent responses. Ultimately, the type of inputs received alters the regulatory network, resulting in constrained cell states, akin to a phenotype being an adaptive product of cellular input. Plausibly, NTS \(Fos^{+}\) cells receive particular combinatorial inputs beyond blood pressure and integrate variable sets of cardiovascular homeostasis modulators such as pain, temperature, exercise, or mood, all of which affect cellular state and input processing. The influence of various inputs on NTS cell states is symbolically represented in Figure 7C (Paton 1998; Dampney and Horisuchi 2003; Michelin 2007; Rinaman 2011; Grill and Hayes 2012). Such input-based influences imply that NTS neurons are individually gated in dynamic responses to combinatorial inputs, rather than behaving as a homogeneous population and integrating all inputs into a population rate code. NTS neurons dynamically...
responding to inputs shape a mechanism of blood pressure homeostasis based on the selection or gating of particular NTS neurons activated by combinatorial demands on blood pressure.

Similarly demonstrating a functional meaning to variability, Marder and Taylor (2011) have shown that variability extends to the levels of electrical and neural network function. In this
mechanism, the “neural code” by which blood pressure regulation is performed would be based on molecular states of individual neurons. This novel explanation of blood pressure homeostasis in the brain supports the emerging perspective reflected in the recently announced BRAIN Initiative. With the convergence of sophisticated experimental techniques and accurate and precise high-throughput technologies we have a unique opportunity to develop an integrated view of molecular identity (DNA sequence, single-cell transcriptomes, epigenomic information, and protein expression). This picture, in combination with information on anatomical connectivity and functional measures (e.g. physiology) will afford an unprecedented view of the vertebrate brain. (NIH RFA-MH-14-215 2013)

This perspective, supported by our results, expands the definition of a neuronal cell type to include post-developmental plasticity and highlights the role of transcriptional regulation in shaping these phenotypes. Furthermore, an alternative hypothesis can now be proposed on how a cell population supports robust biological function: Functional robustness is achieved through the development of a graded set of cellular responses, rather than a uniform population response. Analyses of this type could also be extended outside the brain to other environments where cells clearly vary and receive different inputs. Finally, the identification of cell type specific gene network topologies may be facilitated by the organization of variability in the transcriptional identity and response of individual cells.

**Methods**

We collected 300 single neurons lifted from the NTS of six rats (hypertensive rats $n = 4$, normotensive rats $n = 2$), 220 of which were collected from hypertensive rats and the remaining 80 individual neurons collected from normotensive rats. Due to the absence of a perturbation in arterial blood pressure in normotensive rats, Fos$^+$ cells responding to an acute hypertensive challenge were only collected in hypertensive rats. The anatomical distribution of collected Fos$^+$ neurons within the NTS was consistent with the extensive literature characterizing Fos expression in the NTS (Li and Dampney 1992, 1994; Chan and Sawchenko 1994, 1995, 1998; Miura et al. 1994; Graham et al. 1995; Shih et al. 1996; Chan et al. 1998, 2000). Acute hypertension challenge, immunohistochemistry staining, and LCM are described in greater detail in Supplemental Material and Methods.

![Figure 7](image-url)
High-throughput qPCR
Gene expression levels were measured across four high-throughput qPCR assay chips on the BioMark (Fluidigm), a highly reproducible qPCR platform, which has demonstrated minimal technical variability over five orders of gene expression (Supplemental Fig. S1). Additional details regarding the high-throughput qPCR (Spurgeon et al. 2008) are described in Supplemental Material and Methods.

Data normalization
A total of 300 single cells and 96 gene assays were collected, which were reduced by rigorous quality control (QC) to 192 single cell samples (41 normotensive samples and 151 hypertensive samples) and 81 different gene assays that were included in the present analysis.

Raw Ct values for individual samples were normalized against an average expression level between Actb and Rpl19 to obtain a $-\Delta C_i$ and $-\Delta C_{i}^{R}$ value (Spurgeon et al. 2008) was used in order to relate this value to actual gene expression (e.g., a $-\Delta C_i$ value of 10 in one cell has higher gene expression than a cell with a $-\Delta C_i$ value of 5 or $-2$ for a particular gene). The following equation was used to calculate $-\Delta C_i$:

$$-\Delta C_i^{R} = \text{average}(C_i^{act}, C_i^{Rpl19}) - C_i^{R}.$$ (1)

Actb and Rpl19, included as part of a set of potential housekeeping genes, were selected based on previously developed methods (Vandesompele et al. 2002; Andersen et al. 2004). The $-\Delta C_i$ values were used as a measure for relative gene expression and used as the basis for the analytical methods utilized in this report.

Principal Component Analysis (PCA)
The pcaMethods package (Stacklies et al. 2007) and associated functions in the R statistical software (R Development Core Team 2013) were used to perform PCA. A subset of 48 genes was derived as significantly contributing to the observed variability, using the five highest and lowest corresponding loading values along the first five principal components as a basis for gene identification. For further details please refer to Supplemental Material and Methods.

Gene correlation networks
The statistical software R was used to determine rank correlation coefficients between the subset of 48 genes for the six different single cell “sub-phenotypes” initially identified. A Spearman rank correlation coefficient cutoff of 0.4 was used to define whether or not two genes had a correlative relationship. Cytoscape (www.cytoscape.org) was used to visualize the correlative relationships.

Distance and multidimensional scaling (MDS)
Relative distances between single cells were determined using the Spearman rank correlations obtained for the set of hypertensive samples and subset of 48 genes identified from PCA. The following equations were used to determine distance:

$$\rho = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}.$$ (2)

where $x_i$ and $y_i$ correspond to gene expression rank between two single cell samples;

$$d = 1 - \rho.$$ (3)

where $d$ corresponds to the Spearman rank distance between two cells.

The pairwise relative distances between single cell samples was performed using the stats package provided through the R statistical software (R Development Core Team 2013).

Nonmetric MDS was performed on single cells in conjunction with PCA in order to analyze single cells that lie in an n-dimensional space (due to the nature of the multiplex gene expression data). MDS was performed on the single cells obtained from hypertensive rats since genes were rank-ordered for each sample (based on $-\Delta C_t$). The MDS would then map the relative distances between the samples onto a lower-dimensional plane, while minimizing the error observed between the actual distances in the n-dimensional space and the lower-dimensional space (i.e., minimizing the stress) (Van Deun and Delbeke 2000). The isoMDS function provided in the MASS package (Venables and Ripley 2002) for R platform was used to perform the MDS. Following MDS, single cell samples were plotted in the lower-dimensional MDS space with the input-type marker categorization (combinatorial expression levels of Th and Fos) overlaid on the samples. The first MDS axis discriminates samples based on the rank-ordering of expression levels of genes from transcription module 2. MDS axis 2 accounts for biological variability in both hypertensive and baseline samples (Supplemental Fig. S13), and MDS axis 3 discriminates cells based on rank expression levels of genes from transcription module 1. Both two-dimensional and three-dimensional plots were created via plotrix and rgl packages (Lemon 2006; Adler and Murdoch 2013) provided by the R statistical software (R Development Core Team 2013).

Data access
Both raw Ct and $-\Delta C_i$ values of samples passing quality control are included as Supplemental Material (Supplemental Tables S1, S2).

Acknowledgments
The work presented here is funded through NIH NIGMS R01 GM083108 and NIH NHLBI R01HL116621. J.P. acknowledges Mrs. Sirisha Achanta for guidance and assistance with performing the LCM quality control experiments.

Author contributions: J.P. performed analysis and contributed to writing and figure design; A.B. performed analysis, writing, and figure design; K.K. was involved in design, initial analysis, experimental execution, and analysis of blood pressure data; S.G. contributed to design, single cell LCM, and staining methods; B.O. was involved in initial analysis and editing; J.S. designed the study and was involved in analysis, writing, and figure design; R.V. designed the study and was involved in analysis, writing, and figure design. J.P. and A.B. contributed equally to the study. All authors discussed the results and commented on the manuscript.

References


www.genome.org

940 Genome Research


Received June 7, 2013; accepted in revised form March 25, 2014.