1-1-2015

Single-Cell Gene Expression Variation as A Cell-Type Specific Trait: A Study of Mammalian Gene Expression Using Single-Cell RNA Sequencing

Hannah R. Dueck
University of Pennsylvania, hdueck@mail.med.upenn.edu

Single-Cell Gene Expression Variation as A Cell-Type Specific Trait: A Study of Mammalian Gene Expression Using Single-Cell RNA Sequencing

Abstract
In this dissertation, we used single-cell RNA sequencing data from five mammalian tissues to characterize patterns of gene expression across single cells, transcriptome-wide and in a cell-type-specific manner (Part 1). Additionally, we characterized single-cell RNA sequencing methods as a resource for experimental design and data analysis (Part 2).

Part 1: Differentiation of metazoan cells requires execution of different gene expression programs but recent single cell transcriptome profiling has revealed considerable variation within cells of seemingly identical phenotype. This brings into question the relationship between transcriptome states and cell phenotypes. We used high quality single cell RNA sequencing for 107 single cells from five mammalian tissues, along with 30 control samples, to characterize transcriptome heterogeneity across single cells. We developed methods to filter genes for reliable quantification and to calibrate biological variation. We found evidence that ubiquitous expression across cells may be indicative of critical gene function and that, for a subset of genes, biological variability within each cell type may be regulated in order to perform dynamic functions. We also found evidence that single-cell variability of mouse pyramidal neurons was correlated with that in rats consistent with the hypothesis that levels of variation may be conserved.

Part 2: Many researchers are interested in single-cell RNA sequencing for use in identification and classification of cell types, finding rare cells, and studying single-cell expression variation; however, experimental and analytic methods for single-cell RNA sequencing are young and there is little guidance available for planning experiments and interpreting results. We characterized single-cell RNA sequencing measurements in terms of sensitivity, precision and accuracy through analysis of data generated in a collaborative control project, where known reference RNA was diluted to single-cell levels and amplified using one of three single-cell RNA sequencing protocols. All methods perform comparably overall, but individual methods demonstrate unique strengths and biases. Measurement reliability increased with expression level for all methods and we conservatively estimated measurements to be quantitative at an expression level of ~5-10 molecules.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Genomics & Computational Biology

First Advisor
Junhyong Kim

SINGLE-CELL GENE EXPRESSION VARIATION AS A CELL-TYPE SPECIFIC TRAIT:

A STUDY OF MAMMALIAN GENE EXPRESSION USING SINGLE-CELL RNA SEQUENCING

Hannah R. Dueck

A DISSERTATION

in

Genomics and Computational Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation

_____

Junhyong Kim
Professor of Biology

Graduate Group Chairperson

_____

Li-San Wang, Associate Professor of Pathology and Laboratory Medicine

Dissertation Committee

Arjun Raj, Assistant Professor of Bioengineering, University of Pennsylvania
James Eberwine, Professor of Pharmacology, University of Pennsylvania
Nancy Zhang, Associate Professor of Statistics, University of Pennsylvania
Isidore Rigoutsos, Director of Computational Medicine Center, Thomas Jefferson University

SINGLE-CELL GENE EXPRESSION VARIATION AS A CELL-TYPE SPECIFIC TRAIT:

A STUDY OF MAMMALIAN GENE EXPRESSION USING SINGLE-CELL RNA SEQUENCING

Dedication

For Ryan,

whose encouragement helped me take the first step,
and the next one.

And for Mom and Dad,
Allan and Laura Sue Dueck,

whose lives demonstrate the courage of persistence
and the joy of creativity.

ACKNOWLEDGMENTS

Many people have contributed to my graduate school education and research, and I greatly appreciate each one. Junhyong Kim, my advisor, has provided invaluable instruction, stimulating conversations, and much needed encouragement throughout. I particularly value his thoughtful approach to research, conceptual creativity, attention to experimental and analytic detail, educational lab meetings, and supportive mentorship.

I have learned a lot from discussions with all of the members of the Kim lab, and have enjoyed our shared commiserations and laughter. Particular thanks to Mugdha Khaladkar, Youngji Na, Sarah Middleton and Stephen Fisher, who have been great sounding boards regarding computational analyses, and to Chantal Francis and Hoa Giang, who have provided insight into the workings of a wet lab.

My committee has offered thoughtful suggestions, advice and support: Arjun Raj, Nancy Zhang, Rigoutsos. I am particularly indebted to James Eberwine, who has guided and collaborated with me in all of the research contained in this dissertation.

Much of the work in this dissertation is rooted in collaborative research. Chapter 1 includes material prepared in collaboration with Junhyong Kim and James Eberwine that has been submitted for publication. Work in this Chapter 2 was performed in collaboration with Mugdha Khaladkar, Tae Kyung Kim, Jennifer M. Spaethling, Chantal Francis, Sangita Suresh, Stephen A. Fisher, Patrick Seale, Sheryl G Beck, Tamas Bartfai, Bernhard Kuhn, James Eberwine, and Junhyong Kim and was published in Genome Biology (Dueck et al., 2015a). Material from this publication is presented in the dissertation abstract and in Chapter 2 with limited modification. The research in Chapter 3 was performed in collaboration with Rizi Ai, Ray Dominguez, Oleg Evgrafov, Jian-Bing Fan, Stephen Fisher, Chantal Francis, Jennifer Hernstein, Tai Kyung Kim, Hugo Kim, Sonia Lin, Rui Liu, Bill Mack, Neeraj Salathia, Jennifer Spaethling, Tade Souaiaia, Jai-Yoon Sul, Andre Wilberg, Robert Chow, James Eberwine, James Knowles, Kun Zhang, and Junhyoung Kim, and a manuscript describing this work is in preparation.

Dissertation research was funded in part by a National Science Foundation Graduate Research Fellowship, and by a fellowship under the NIH Integrated Interdisciplinary Training Program in Computational Neuroscience at the University of Pennsylvania (T90 DA022763-05).

And thank you, especially, to my family – for sending encouragement when it was needed, celebrating milestones, and sharing life with me. And to Ryan, my gratitude is beyond words.

ABSTRACT

SINGLE-CELL GENE EXPRESSION VARIATION AS A CELL-TYPE SPECIFIC TRAIT:
A STUDY OF MAMMALIAN GENE EXPRESSION USING SINGLE-CELL RNA SEQUENCING

Hannah R. Dueck

Junhyong Kim

In this dissertation, we used single-cell RNA sequencing data from five mammalian tissues to characterize patterns of gene expression across single cells, transcriptome-wide and in a cell-type-specific manner (Part 1). Additionally, we characterized single-cell RNA sequencing methods as a resource for experimental design and data analysis (Part 2).

Part 1: Differentiation of metazoan cells requires execution of different gene expression programs but recent single cell transcriptome profiling has revealed considerable variation within cells of seemingly identical phenotype. This brings into question the relationship between transcriptome states and cell phenotypes. We used high quality single cell RNA sequencing for 107 single cells from five mammalian tissues, along with 30 control samples, to characterize transcriptome heterogeneity across single cells. We developed methods to filter genes for reliable quantification and to calibrate biological variation. We found evidence that ubiquitous expression across cells may be indicative of critical gene function and that, for a subset of genes, biological variability within each cell type may be regulated in order to perform dynamic functions. We also found evidence that single-cell variability of mouse pyramidal neurons was correlated with that in rats consistent with the hypothesis that levels of variation may be conserved.

Part 2: Many researchers are interested in single-cell RNA sequencing for use in identification and classification of cell types, finding rare cells, and studying single-cell expression variation; however, experimental and analytic methods for single-cell RNA sequencing are young and there is little guidance available for planning experiments and interpreting results. We characterized single-cell RNA sequencing measurements in terms of sensitivity, precision and accuracy through

analysis of data generated in a collaborative control project, where known reference RNA was diluted to single-cell levels and amplified using one of three single-cell RNA sequencing protocols. All methods perform comparably overall, but individual methods demonstrate unique strengths and biases. Measurement reliability increased with expression level for all methods and we conservatively estimated measurements to be quantitative at an expression level of ~5-10 molecules.

TABLE OF CONTENTS

ACKNOWLEDGMENTS IV

ABSTRACT VI

LIST OF TABLES XII

LIST OF ILLUSTRATIONS XIV

CHAPTER 1: THEORIES OF SINGLE-CELL TRANSCRIPTOME HETEROGENEITY 1

1.1. Introduction 1

1.2. Arrival of single-cell RNA sequencing 2

1.3. Prevalent theories of single-cell transcriptome heterogeneity 4

1.3.1. Cell type diversity 4

1.3.2. Asynchronous dynamic processes 5

1.3.3. Molecular mechanics of gene expression 6

1.3.4. Microenvironmental context 7

1.3.5. Degeneracy, homeostasis and robustness 7

1.4. Hypothesis: Population function may depend on single-cell variation 8

1.4.1. Bet hedging: A pre-existing diversity of cell states allows rapid population adaptation to a new environment. 10

1.4.2. Generalized bet hedging: Random phenotype generation enables population response to novel environments. 12

1.4.3. Response distribution: Variation across single cells may allow a graded population response. 13

1.4.4. Priming and fate plasticity: Gene expression variation endows cells and populations with fate plasticity. 13

1.4.5. Information propagation: Population diversity may enable information coding and transfer. 14

1.4.6. Crowd control: Rare cells respond rapidly to perturbations and coordinate population behavior. 15

1.5. Regulation of expression variation 16

1.6. Discussion 17

1.7. Introduction to dissertation research 19

CHAPTER 2: CELL-TYPE SPECIFIC PATTERNS OF SINGLE CELL TRANSCRIPTOME VARIATION 21

2.1. Background 21

2.2. Results 22

2.2.1. Single-cell RNA-sequencing datasets 22

2.2.2. Single-cell transcriptome complexity 26

2.2.3. Consistent gene expression across single cells 27

2.2.4. Single-cell transcriptome variation 30

2.2.5. Within-cell-type variability 33

2.2.6. Patterns of expression variation 35

2.2.7. Association of extreme variability with RNA half-life 37

2.3. Discussion 37

2.4. Materials and methods 39

2.4.1. Cell culture and single cell isolation 39

2.4.2. mRNA amplification and library construction 40

2.4.3. Dilution controls 40

2.4.4. Alignment, quantification and sample selection 40

2.4.5. Characterization of single-cell transcriptomes 42

2.4.6. Consistent genes 43

2.4.7. Gene expression variability 44

CHAPTER 3: ASSESSMENT OF SINGLE-CELL RNA SEQUENCING METHODS 49

3.1. Overview 49

3.2. Introduction 49

3.3. Results 50
3.3.1. RNA-sequencing datasets 50
3.3.2. Data processing 52
3.3.3. Gene detection sensitivity 54
3.3.4. Precision 58
3.3.5. Accuracy 62
3.3.6. Protocol variations 65

3.4. Discussion 68

3.5. Methods 70
3.5.1. Experimental design 70
3.5.2. Alignment and quantification 72
3.5.3. Excluded and unambiguous genes 75
3.5.4. Expected number of molecules in diluted replicates 75
3.5.5. Genomic distribution of sequenced reads 76
3.5.6. Number of detected genes 77
3.5.7. Gene traits 78
3.5.8. Detection logistic regression 78
3.5.9. Sensitivity outliers 82
3.5.10. Coverage 82
3.5.11. Precision 83
3.5.12. Accuracy 85
3.5.13. Protocol variations 86

CHAPTER 4: CONCLUSION 88

4.1. Introduction 88

4.2. Single-cell RNA sequencing assessment 88
4.2.1. Overall assessment of measurements 88
4.2.2. Biases 89
4.2.3. Measurement reliability and expression level 90
4.2.4. Depth of sequencing and sensitivity 90

4.2.5. Protocol optimizations 91

4.3. Survey of transcriptome heterogeneity across single cells from diverse mammalian tissues 91

4.3.1. Tissue-specific transcriptome characteristics 91

4.3.2. Patterns of single-cell expression variation 92

4.3.3. Evidence suggesting control over the extent of expression variation 92

4.4. Future outlook 93

APPENDIX A: SUPPLEMENTAL MATERIAL FOR CHAPTER 2 96

A.1. Supplemental tables 96

A.2. Supplemental figures 97

APPENDIX B: SUPPLEMENTAL MATERIAL FOR CHAPTER 3 104

B.1. Supplemental tables 104

B.2. Supplemental figures 105

BIBLIOGRAPHY 108

LIST OF TABLES

Table 1.1 Scenarios where aggregate function may depend on single-cell variation. 10

Table 2.1 Dataset and RNA sequencing statistics for each cell type 23

Table 2.2 Gene expression variability of different functional gene categories 34

Table 3.1 Coverage selectivity by method 54

Table 3.2 Evaluation of protocol variations 66

Table A.1 RNA sequencing and quality control data for each single-cell sample 96

Table A.2 Summary of single-cell transcriptome characteristics 96

Table A.3 Identified gene sets 96

Table A.4 Association between common expression and mutant phenotypes 96

Table A.5 Within- and between-cell type transcriptome variance 96

Table A.6 Enrichment of Gene Ontology (GO) categories among genes identified by expression pattern 97

Table A.7 Mammalian Phenotype Ontology annotations used in mutation analysis 97

Table B.1 Control dataset sample identification, protocol information, and RNA sequencing stats 104

Table B.2 Optimization dataset sample identification, protocol information, and RNA sequencing stats 104

Table B.3 Gene detection logistic regression model 104

Table B.4 Gene detection logistic regression fit and validation 104

Table B.5 Probability of gene recovery 104

Table B.6 Gene detection outliers 105

Table B.7 Precision outliers 105

Table B.8 Accuracy outliers 105

LIST OF ILLUSTRATIONS

Figure 2.1 Single-cell dataset and transcriptome characteristics. 25

Figure 2.2 Phenotypic importance of genes with consistent expression. 28

Figure 2.3 Cell-type patterns of transcriptome variability 32

Figure 3.1 Experimental design and sequenced data 51

Figure 3.2 Single-cell RNA sequencing sensitivity 55

Figure 3.3 Single-cell RNA-sequencing precision 60

Figure 3.4: Single-cell RNA sequencing accuracy 63

Figure A.1 Dataset quality and transcriptome characteristics 99

Figure A.2 Accounting for technical characteristics of aRNA sequencing. 101

Figure A.3 A subset of genes demonstrates variable expression in each examined tissue type. 103

Figure B.1 Accuracy and robustness of estimated reference HBR and UHR RNA expression levels 105

Figure B.2 Normalized read counts and expectation for ERCC transcripts 107

CHAPTER 1: Theories of single-cell transcriptome heterogeneity

1.1. Introduction

To understand mammalian tissue function, both in health and in disease, it is important to understand gene expression and gene regulation in a tissue-specific manner. These topics are often addressed by measuring average gene expression values across thousands of cells, under the implicit assumption that individual cells within a tissue or of a common type demonstrate homogenous gene expression profiles; however, recent technological advancements have enabled increasingly high-resolution measurements of gene expression in single cells (Buckley et al., 2011; Hashimshony et al., 2012; Islam et al., 2011, 2014; Jaitin et al., 2014; Lee et al., 2014b; Picelli et al., 2013; Ramsköld et al., 2012; Sasagawa et al., 2013; Tang et al., 2009) resulting in a growing appreciation for the extent of expression variability across cells (Brennecke et al., 2013; Chiu et al., 2014; Deng et al., 2014; Eckersley-Maslin et al., 2014; Park et al., 2014; Piras et al., 2014; Poulin et al., 2014; Sul et al., 2009). This variability has been examined from many vantage points: as an indicator of the vast diversity of cell types in multicellular organisms, or as a byproduct of redundancy in regulatory networks, a temporal snapshot of asynchronous dynamic processes or the product of molecular dynamics, a consequence of microenvironment diversity or as evidence that RNA abundance may be irrelevant for cell phenotype. An alternative perspective is to consider whether single cell transcriptome, proteome, and other molecular variability might be critical for tissue/population-level function. Under this hypothesis, molecular variation indicates a diversity of hidden functional capacities within an ensemble of “identical” cells and this functional diversity facilitates collective behavior that would be inaccessible to a homogenous population.

In this chapter, we briefly introduce single-cell RNA sequencing and survey the prevalent theories found in studies of single-cell variation, listed above. We then turn to an extended exploration of possible population-level functions that depend on heterogeneity of constituent

cells, along with summaries of supporting evidence from recent literature. After a brief summary of recent research on the regulation of gene expression variation, the chapter closes with an introduction to the work comprised in this dissertation.

Before we begin, a comment about pronoun use: Each of the chapters represents works for which I have been the first author and the primary contributor to the analytical and scientific approaches; however, the projects presented here also involved the collaborative efforts of many co-authors. Throughout this dissertation I use to pronoun "we" to indicate the contributions from other collaborators, which is documented in the respective papers.

1.2. Arrival of single-cell RNA sequencing

The advent of gene expression microarrays and RNA sequencing revolutionized studies of gene expression by providing high-resolution measurements transcriptome-wide (Majewski and Pastinen, 2011; Mortazavi et al., 2008; Ozsolak and Milos, 2011; Ponting and Belgard, 2010; Wang et al., 2009). In the 1990s, the gene expression microarray was developed and became a heavily used tool. Typically in this method, fluorescently labeled complementary DNA (cDNA) molecules are hybridized to arrays of oligonucleotide probes and imaged to estimate gene abundance by fluorescence intensity. Expression measurements are high-throughput, but cover a limited detection range bounded by cross-hybridization noise and florescence signal saturation (Mortazavi et al., 2008; Wang et al., 2009). In 2008, Mortazavi et al. developed RNA sequencing, which offers several technological advances. Here, cDNA molecules are fragmented and read using next-generation sequencing techniques. Expression measurements are genome-wide, unbiased by gene models, accurate and digital, capable of detecting alternate poly-adenylation, RNA editing events, non-coding RNA and differential allele expression. Both techniques have been used to uncover highly complex gene expression patterns in mammalian tissues and to characterize the expression heterogeneity that underlies phenotypic variation across tissues and cell types.

Our understanding of the relationship between cell identity and gene regulation has been heavily informed by these population-level expression measurements; however, single cell experiments have shown that gene expression in individual cells may differ substantially from the population average (Levsky and Singer, 2003). Foundational studies, using methods such as targeted gene amplification (Miyashiro et al., 1994) and single molecule fluorescent *in situ* hybridization (smFISH; Raj et al., 2008), demonstrated large heterogeneity in gene expression across cells of the same cell type, and suggested more fine-scaled models of cell identity and transcriptional mechanics (Eberwine and Crino, 2001; Eldar and Elowitz, 2010; Marder and Goaillard, 2006; Munsky et al., 2012; Raj et al., 2008; Raser and O'Shea, 2005). These methods were limited to the measurement of a small number of genes and so could not be used to asses the heterogeneity of gene expression across single cells transcriptome-wide or the relationship of this heterogeneity to single-cell phenotypic variation.

The current available microarray and RNA sequencing methods require substantial amounts of total input RNA (~nanograms to micrograms), orders of magnitude more material than found in a single cell (~femtograms). In 1990, two methods were presented to amplify the whole transcriptome of a single cell to detectable quantities, one using linear amplification (Van Gelder et al., 1990) and another using PCR (Brady et al., 1990; reviewed in Tang et al., 2011). In the early 2000s, these methods were adapted to generate sufficient quantities of cDNA for measurement of single-cell transcriptomes on gene expression microarrays (Kamme et al., 2003; Tietjen et al., 2003) and, in 2009, Tang et al. extended this approach to perform bulk RNA sequencing of PCR-amplified cDNA using next-generation sequencing technologies.

Since this initial single-cell RNA sequencing publication, there has been substantial methods development with nearly all protocols following the same three steps: capture of RNA from a single cell and conversion to cDNA, amplification of cDNA to detectable quantities, and measurement by bulk RNA sequencing. Compared to bulk RNA protocols, this procedure

requires additional sample handling and enzymatic reactions, coupled with substantial molecular amplification, and so there is potential for increased experimental variation. J. Eberwine was a pioneer in recognizing this challenge of single-cell transcriptomics, and his laboratories have optimized enzymatic reactions for small input amounts and developed the linear amplification protocol referenced above, which uses transcription rather than PCR for amplification to limit the exponential expansion of early technical errors (Miyashiro et al., 1994; Van Gelder et al., 1990). Single-cell RNA sequencing methods development has continued to propose novel techniques to limit experimental noise, increase sensitivity, decrease preparation time, increase throughput, multiplex with other –omics measurements, and retain information about RNA spatial position (Buckley et al., 2011; Chen et al., 2015; Dey et al., 2015a; Hashimshony et al., 2012; Islam et al., 2014; Jaitin et al., 2014; Lee et al., 2014b; Lovatt et al., 2014; Picelli et al., 2013; Sasagawa et al., 2013). For further details on experimental methods, see Chapter 3.

1.3. Prevalent theories of single-cell transcriptome heterogeneity

Single-cell RNA sequencing enabled high resolution, multi-genic RNA expression measurements in single cells and studies using this technique have reported extensive expression variation transcriptome-wide. A large body of research has developed in the study of single cell gene expression variation using high-throughput gene expression measurements, and this research has employed a variety of conceptual frameworks for the study of this variation. A brief survey of several prevalent theories follows, in which expression variation is considered to reflect extensive cell type diversity, asynchronous dynamic processes, molecular mechanics of transcription, diverse microenvironmental contexts, or functional degeneracy and phenotypic robustness.

1.3.1. Cell type diversity

Measurements of single-cell transcriptomes allow a high-resolution measurement of cell identity and the opportunity to discover rare cell types (Chiu et al., 2014; Grun et al., 2015; Jaitin

et al., 2014; Poulin et al., 2014; Treutlein et al., 2014; Usoskin et al., 2015). Under the assumption that (discrete) functional diversity is reflected in (discrete) molecular diversity, gene expression variation indicates the presence of an assortment of cell types and sub-types. Single-cell RNA sequencing data may be used to identify clusters of similar molecular states and rare outlier states, which can then be categorized as cell types. Distinctive molecular signatures of each cluster may be used as a high resolution, unbiased and quantitative cell type classification criteria, informative with regards to cell physiology and function. Because tissue function depends on the functions and interactions of its constitutive cells, cataloguing cell types may advance our understanding of tissue behavior. This framework has been used to identify novel or rare cell types, diverse cell types in complex tissues, and transient cell types (Chiu et al., 2014; Grun et al., 2015; Jaitin et al., 2014; Poulin et al., 2014; Treutlein et al., 2014; Usoskin et al., 2015). In a prototypical example, Trutlein et al. sequenced dissociated single cells from the distal lung epithelium at four different developmental time points (Treutlein et al., 2014). By clustering these gene expression profiles, the authors identified five distinct cell types, four assigned to known classes based on the expression of previously annotated marker genes, and a fifth hypothesized to be a (transient) bipotential progenitor.

1.3.2. Asynchronous dynamic processes

Normal cell function requires dynamic changes in cell state through processes such as the cell cycle, circadian rhythm or differentiation, and these changes are manifest in temporal variation of gene expression (Bieler et al., 2014; Buettner et al., 2015; Durruthy-Durruthy et al., 2014; McDavid et al., 2014; Moignard et al., 2015; Treutlein et al., 2014). Even if these processes progress in a uniform manner for all cells, differences in observation time or asynchrony across a population will be apparent as expression variation in single-cell RNA sequencing data. In this case, expression variation across single cells may be used to infer a temporal ordering of cells or gene expression states, and expression co-variation may be used to infer dynamic regulatory processes. For example, on the discovery of hypothesized bipotential progenitor cells in the distal

lung, Trutlein et al. identified marker genes that distinguished these progenitors from mature cell types and used the expression level of these markers in single cells to pseudo-order cells by stage in dynamic differentiation process. Using this ordering, the authors inferred dynamic expression patterns that occur over development of the alveolar lineages.

1.3.3. Molecular mechanics of gene expression

The reactions underlying gene expression involve small numbers of molecules, with many genes encoded on only two molecules of DNA. Observed gene expression variation may result from the stochastic nature of these molecular reactions and, conversely, gene expression variation may be used to study the mechanistic process of gene expression. Foundational studies demonstrated the occurrence random expression variation within single cells (Elowitz et al., 2002; Raj et al., 2006) and developed probabilistic models to describe the process of gene expression (reviewed in Munsky et al., 2012; Paulsson, 2005). Two primary models of gene expression were established: the constitutive model, parameterized by rates of transcription and degradation, and the two-state model, which incorporates dynamic changes in the gene accessibility for transcription and is additionally characterized by the rate at which gene accessibility switches (the "burst frequency") and the number of RNA molecules that are transcribed during a window of accessibility (the "burst size") (Raj et al., 2006). Recent studies have used single-cell RNA sequencing to infer model parameters and to study transcription kinetics across many genes and across experimental conditions (Grün et al., 2014a; Kim and Marioni, 2013; Piras et al., 2014). In a pioneering study, Grün et al. modeled experimental variation in single-cell RNA sequencing and, using this model, inferred the extent of biological noise for highly expressed genes in mouse embryonic stem cells (ESCs) grown in traditional serum and in 2i conditions, as well as parameters for the two-state gene expression model for these genes. Based on these data, the authors found larger expression variation among ESCs grown in serum and showed that this effect could be explained by larger burst size but lower burst frequency.

1.3.4. Microenvironmental context

Each cell in a population likely experiences and adapts to a unique micro-environmental and social context, even in environments that are assumed to be homogenous (Snijder and Pelkmans, 2011). Single-cell expression profiles may reflect regulatory states shaped by the local microenvironment, and expression variation across cells may indicate underlying environmental and social diversity. Several recent expression profiling studies have considered the contribution of spatial context to the diversity of expression across cells[1], and in several cases have demonstrated that single cell gene expression levels reflect both cell type and cell location or microenvironment (Durruthy-Durruthy et al., 2014; Lovatt et al., 2014; Park et al., 2014; Patel et al., 2014; Poulin et al., 2014; Satija et al., 2015). This relationship between cell diversity and spatial context may be of particular interest in cancerous tumors, which can contain large spatial diversity in nutrient and growth factor availability (Patel et al., 2014) and in development, where morphogen gradients and local contacts direct cell fate decisions (Durruthy-Durruthy et al., 2014; Ohnishi et al., 2014; Satija et al., 2015).

1.3.5. Degeneracy, homeostasis and robustness

A dominant model for generation of cell phenotype suggests that multi-genic phenotypes depend on specific expression levels of underlying genes; however, it may be possible for stable phenotypes to be generated by a diversity of expression profiles (Kim and Eberwine, 2010; Kim et al., 2011; Marder and Goaillard, 2006; Moignard et al., 2015; O'Leary et al., 2013; Park et al., 2014; Rifkin et al., 2000; Schulz et al., 2007; Sul et al., 2009). For example, neurons may generate similar net ionic currents by way of many different combinations of composite ion

---

[1] Typically, single cells are isolated by cell dissociation before preparation of single-cell RNA sequencing libraries, and so spatial information is lost. Several novel methods have been developed in order to assign spatial positions and multi-genic expression measurements to individual cells (Chen et al., 2015; Durruthy-Durruthy et al., 2014; Lee et al., 2014b; Lovatt et al., 2014; Satija et al., 2015).

channels. Under this framework, observed expression variation across homogenous cells indicates the space of expression profiles that generate the same stable phenotype, and, conversely, multi-genic constraints on expression variation (as in expression co-variation) indicate the boundaries of permissible profiles. This “molecular equi-phenotype set” (Kim and Eberwine, 2010) may reflect homeostatic regulation, or control of gene expression levels to maintain a stable phenotype despite dynamic changes in context (such as changes in cell size) (O’Leary et al., 2013; Padovan-Merhar et al., 2015). Alternatively, this diversity of permissible profiles may reflect gene functional degeneracy or redundancy in regulatory networks, such that expression levels for individual genes need not be tightly regulated as long as multi-genic expression levels remain within some bounds (Kim and Eberwine, 2010; Raj et al., 2010; Rifkin et al., 2000; Schulz et al., 2007). In either case, expression variation is reflective of phenotype robustness to perturbations, and the bounds on expression variation contain information about the regulatory system underlying the stable cell phenotype. From this perspective, cell type might be best classified by patterns of expression co-variation among cells with homogenous phenotypes, rather than by specific expression levels of marker gene expression (Kim and Eberwine, 2010; O’Leary et al., 2013).

1.4. Hypothesis: Population function may depend on single-cell variation

In much of the literature discussed above, single-cell heterogeneity is considered at the level of individual cells, with a focus on intracellular gene regulation and transcription and on the relationship between gene expression and cellular phenotype. An interesting alternative perspective is to consider expression variation as a characteristic of the population and with respect to population/tissue phenotype. Are there scenarios where single cell transcriptome variability might be critical for tissue/population-level function? In this section, we review recent literature suggesting that, in some cases, aggregate population-level behaviors may stem from

single-cell heterogeneity. Material in this section has been submitted for publication (Dueck et al., 2015a) and is presented below with modifications.

Before we begin, it will be valuable to consider a few preliminary details. Though literature on gene expression variation segregates into several sub-areas, as seen above, a number of terms are used interchangeably. In this section, we use “single-cell variation” or “single-cell heterogeneity” in reference to diversity among cells presumed to be of the same well-recognized cell type, and not in reference to a diversity of cell types. Second, as discussed above, single-cell expression variation may reflect homeostatic mechanisms and the robustness of cell phenotype to perturbations. It is well established that many population-level processes critically depend on the stability of cellular phenotype, as in development or neural circuit function (Marder and Goaillard, 2006; Raj et al., 2010). Here, we restrict our focus to the idea that variation, in- and of- itself, is required for population behavior. Third, in this section, the phrase “stochastic gene expression” is used in reference to the probabilistic nature of transcription. Use of terms such as “noise”, “random”, “probabilistic” or “stochastic” does not imply a lack of regulation or of utility—gene expression may have randomness but the characteristics of the resulting variation may be regulated. The ability to control the extent of expression variation may be critical if this variation is functional at the population level. More generally, the focus of the following discussion is on population-level behaviors/phenotypes that may stem from single-cell expression variation, and so it is somewhat agnostic to the root causes of this variability. Finally, “function” is a slippery term with many possible scales of reference, from biochemical activity to evolutionary fitness. Here, we do not attempt a precise definition of this term and use it loosely. Regardless of its precise definition, this conversation focuses on how variation among individual cells might, in aggregate and precisely because there is variation, generate higher-level population behaviors / functions.

We now turn to a review of recent literature suggesting aggregate population-level behaviors that may critically depend on cell-to-cell heterogeneity. This review is organized by hypothesized population behavior (see Table 1.1).

Table 1.1 Scenarios where aggregate function may depend on single-cell variation.

Hypothesis	**Description**
Bet hedging	A pre-existing diversity of cell states allows rapid population adaptation to an unpredictable environmental change.
Generalized bet hedging	Extensive randomized phenotypic diversity allows population adaptation of vast diversity of environments.
Response distribution	Cell-to-cell variation in binary decisions allows a fractional or dose-dependent population response.
Fate plasticity and priming	Uncorrelated, sub-threshold fluctuations in regulators of cell fates create subpopulations of cells primed for multiple fate decisions.
Information coding and transfer	A diverse ensemble of individuals enables the population to encode and transmit complex information.
Crowd control	Rare individuals with capacity to respond to perturbation emit local signals that coordinate population behavior.

1.4.1. Bet hedging: A pre-existing diversity of cell states allows rapid population adaptation to a new environment.

In fluctuating, unpredictable environments, a population may benefit by maintaining a diversity of cell phenotypes, each advantageous in a distinct context. This cellular heterogeneity increases the likelihood that the population will contain a subset that is well suited to an unforeseeable future environment. Unlike a strategy where individual cells sense and respond to the environment, maintenance of a standing diversity may be preferred when a rapid response of at least a subpopulation is advantageous and there is insufficient time for signal transduction (Acar et al., 2008). Because this maintenance of diversity protects against a future crisis, this behavior has been termed “bet hedging” and has been extensively studied in single-celled

organisms (Acar et al., 2008; Beaumont et al., 2009; Eldar and Elowitz, 2010; Losick and Desplan, 2008; Martins and Locke, 2015; Mineta et al., 2015). In a classic example, *E. coli* populations maintain a subset of cells in persistence, a quiescent phenotypic state (Eldar and Elowitz, 2010; Losick and Desplan, 2008). Though the presence of persistent cells reduces population growth in nutrient-rich environments, it allows the population to survive unexpected antibiotic agents that target rapidly proliferating cells. To generate the standing population diversity in a uniform environment, individual *E. coli* cells stochastically switch into and out of persistence. Because of this, when an environment conducive to growth returns, the population diversity returns and population growth continues. Bet hedging relies on this type of phenotype switching, where single cells randomly transition between stable phenotypic states. Phenotype switching has been observed broadly, suggesting that this single cell behavior provides a fitness advantage in certain contexts (Losick and Desplan, 2008). Experimental evolution of *Pseudomonas fluorescens* demonstrated that, under a fluctuating selection regime, stochastic phenotype switching could evolve (Beaumont et al., 2009). The rate of bi-stable state switching can be a function of the gene regulatory network and can affect fitness, with an optimal switching rate dependent on the rate of environmental fluctuations (Acar et al., 2008).

We know of no cases of bet hedging in healthy mammalian tissues, perhaps because of the interdependence of cells in multicellular organisms (Losick and Desplan, 2008) or lack of experiments assessing individual cell turnover dynamics; however, it may be that mammalian cancers exhibit this behavior (Gupta et al., 2011; Lee et al., 2014a; Zhou et al., 2014). As in the *E. coli* example, cancer populations may survive chemotherapies that target proliferating cells by switching into and out of a proliferative state, and some studies suggest that such switching occurs (Gupta et al., 2011; Zhou et al., 2014). Phenotype switching has also been hypothesized to play a role in cancer metastasis. Lee et al. characterized a regulatory network that may be capable of producing coexisting noninvasive and pro-metastatic expression states within a triple-negative breast cancer population (Lee et al., 2014a). Models of this regulatory network

suggested that transient perturbations could trigger a cancer cell to switch into a malignant state, potentially initiating metastasis, and that pro-metastatic cells may relax back into a noninvasive state, potentially facilitating the formation of secondary tumors. The implication for functional relevance in these cases is only speculative, and the key question is whether normal cells might employ such bet-hedging strategies. One obvious possibility is with tissues such as skin that directly interact with unpredictable external environment or unpredictable changes in whole organism physiology (e.g., injury response). A more speculative possibility is in developmental contexts where cell proliferation and death in response to patterning gradients is part of morphogenesis.

1.4.2. Generalized bet hedging: Random phenotype generation enables population response to novel environments.

If the diversity of environments that may be encountered is vast, it may be of use for a population of cells to contain as broad a range of phenotypes as possible—to have individuals extensively sample phenotypic space, potentially through use of random mechanisms such as highly variable transcription, errors in transcription or DNA replication, or random genomic rearrangements (Briney and Jr, 2013; Erwin et al., 2014; Jaeger et al., 2013; Woods, 2014). We may consider this as a more generalized form of bet hedging. Though under this strategy individual phenotypes may not be reproducible, it may be that the population benefits substantially by containing at least one successful phenotype. This strategy is observed in multicellular organisms, with archetypal examples including the adaptive immune system, where the diversity of individuals carrying unique genetic re-arrangements provides protection against unknown pathogens (Briney and Jr, 2013; Jaeger et al., 2013), and the stress response triggered by lethal danger, where the generation of diversity through increased molecular error rates may produce an individual who survives (Lee et al., 2014c). The benefits of such extensive diversity may also be relevant in disease. Cancer populations are highly heterogeneous, molecularly and phenotypically, and this population heterogeneity has been associated with resistance to drug

treatment and patient survival (Kim et al., 2015; Lee et al., 2014c; Patel et al., 2014; Singh et al., 2010).

1.4.3. Response distribution: Variation across single cells may allow a graded population response.

Tissues rely on binary decisions made by individual cells, such as whether to enter the cell cycle or apoptosis. Uniformity across cells in binary decisions would produce switch-like population behavior, and in many cases this would be undesirable. Instead, fractional quantitative responses can be achieved by integrating expression fluctuations in decision-making, fluctuations that may be generated by stochastic gene expression. Recent studies have suggested this type of heterogeneity-dependent population behavior in contexts such as fractional population death in response to chemotherapy (Kim et al., 2015; Spencer et al., 2009), maintenance of adult adipose tissue size by fractional differentiation of pre-adipocytes (Ahrends et al., 2014), and graded response to growth factors in the decision to enter the cell cycle in mammary epithelial cells (Overton et al., 2014).

1.4.4. Priming and fate plasticity: Gene expression variation endows cells and populations with fate plasticity.

For some tissues, function depends on cell fate plasticity or the ability of its members to take on a diversity of cell states, as in stem cell populations. Fate plasticity has been associated with expression variation such as stochastic, semi-indiscriminant gene activation, which has also been called “promiscuous gene expression” (Chang et al., 2008; Kumar et al., 2014; Sansom et al., 2014). Numerous studies of stem and progenitor cells have reported variable expression of developmental regulators and have associated this with heterogeneity in differentiation potential (Chang et al., 2008; Islam et al., 2011; Kumar et al., 2014; Ohnishi et al., 2014). Recently, Kumar et al. reported extensive variation across mouse pluripotent stem cells (PSCs) in the activation of stem- and cell-fate regulators, as well as genes that sense and respond to environmental cues,

and also that the extent of variation in the population was associated with the rate of differentiation. In an environment containing cues for both differentiation and also self-renewal, the authors found multiple subpopulations of PSCs: one subpopulation demonstrated relatively homogenous expression and a bias towards self-renewal; a second showed variable activation of cell-fate regulators and also higher rates of spontaneous differentiation.

In many cases, genes critical to cell fate decisions are involved in regulatory networks with switch-like behavior. As gene expression levels approach the network's switching threshold, the probability that induction by external cues will trigger threshold crossing is increased. A cell with expression level near a threshold level for phenotypic switching might be considered to be "primed" for a cell-fate decision (Chang et al., 2008; Kumar et al., 2014). If the population contains a set of cells at variable distance from the threshold, a subset of cells might be always ready to immediately cross the threshold. If the expression state of any individual fluctuates over time, as seen in populations of pluripotent stem cells, then, even as cells differentiate, the population may maintain a characteristic diversity of primed cells (Chang et al., 2008; Kumar et al., 2014).

1.4.5. Information propagation: Population diversity may enable information coding and transfer.

There is an association of high variation with high information content (i.e., high entropy). Single-cell variation can represent both high information content and, if cells are processing information, the capacity to transfer high information content. For example, medullary thymic epithelial cells (mTEC) stochastically transcribe tissue-restricted genes in the mTEC population so that collectively the population exposes thousands of self-antigens to developing T cells. This diversity plays an instructive role in T cell differentiation, so that only T cells with low self-affinity are directed to an effector fate (Hogquist et al., 2005). In the brain, extensive phenotypic diversity may broaden the extent of possible neural circuitry and so enhance the brains capacity for information transfer (Erwin et al., 2014; Muotri et al., 2005, 2010; O'Rourke et al., 2012).

Increased rates of Line1 (L1) retrotranposition, a source of somatic genetic diversity, have been found during neurogenesis, speculatively supporting a functional advantage to heightened diversity in the brain (Muotri et al., 2005).

1.4.6. Crowd control: Rare cells respond rapidly to perturbations and coordinate population behavior.

Several recent single-cell studies of anti-viral or inflammatory response and cell fate choice have reported on cases where a rare subset of cells in a population responded rapidly to perturbation and emitted signals that coordinated population behavior (Fang et al., 2013; Patil et al., 2015; Rand et al., 2012; Shalek et al., 2014; Xue et al., 2015). Described as sentinels, first responders, precocious cells and pioneers, these cells uniquely expressed (Fang et al., 2013; Patil et al., 2015; Rand et al., 2012; Shalek et al., 2014) and secreted (Xue et al., 2015) key cytokines in response to the stimulus. By contrast, the majority of the population was incapable of responding in kind to the same stimulus, even over extended periods of time (Patil et al., 2015; Rand et al., 2012; Shalek et al., 2014; Xue et al., 2015). This two-tiered signaling mechanism coordinated population behaviors, eliciting a uniform response or more complex behavior, such as modulating phenotype heterogeneity spatially or over time. Single-cell variation might encompass the hidden differentiated role of individual cells in control of dynamic whole population behavior.

The idea of sentinel or first responder cells is that a subset of cells in a signal responding state can dynamically reprogram the greater cell population and this helps balance competing needs of the physiological dynamics. Such competing demands might be observed in the immune system, which requires a balance between rapid response to assault and avoidance of self-toxicity (Patil et al., 2015; Shalek et al., 2014). Recently, Patil et al. reported that when human dendritic cells were infected with Newcastle disease virus, a small fraction of cells activated Ifnb1 promptly. Paracrine signals emitted by these early responders activated Ifnb1 expression in the

majority of cells, but in a manner that elicited large variation across cells in time to activation. Population behavior was context dependent, suggesting regulation of the population-level response. Dynamic coordinated population behavior activated through single-cell variation may also be critical in other contexts, such as tissue morphogenesis. A recent study provided suggestive evidence, reporting that a subpopulation of rare cells was essential in normal breast epithelial cell morphogenesis in 3D culture for enforcement of quiescence (Bajikar et al., 2014).

1.5. Regulation of expression variation

Evidence that the extent of expression variation is regulated provides suggestive evidence that variation has functional importance (though see further commentary below). There is an increasing number of studies demonstrating that the distribution of gene expression across cells can be modulated by regulation, through mechanisms involving promoter accessibility, transcript degradation rate, gene copy number, or regulatory network structure (Ahrends et al., 2014; Benayoun et al., 2014; Dar et al., 2014; Dey et al., 2015a, 2015b; Kumar et al., 2014; Lagha et al., 2013). Additionally, multiple studies have suggested separate control of expression mean and variance (Benayoun et al., 2014; Dar et al., 2014; Dey et al., 2015b). Recently, Lagha et al. showed that paused Pol II decreased temporal variation in gene activation in response to Dpp signaling during Drosophila development (Lagha et al., 2013). Through a series of transgene experiments, the authors found that promoter proximal regions regulated the extent of Pol II pausing and that, in turn, the extent of Pol II pausing was correlated with time to synchrony across cells (the time until 50% of cells that ultimately express a gene have activated transcription of that gene). By interchanging transgene promoters, the authors demonstrated that promoter sequence could modulate the extent of expression variation across cells in the developing drosophila embryo, as well as qualitatively change the pattern of gene expression across the population, generating bimodal gene expression in place of synchronous behavior. Benayoun et al. provided evidence that genes broadly covered with H3K4me3 histone modifications exhibit low

expression variation, uncorrelated with expression level (Benayoun et al., 2014). Experimental modification of the breadth of H3K4me3 domains decreased expression consistency, and remodeling of H3K4me3 domains for some genes was observed over adipocyte development, suggesting active regulation of the extent of gene expression consistency over development. Genes with broad H3K4me3 marks associated with Pol II pausing and elongation, suggesting a transcriptional mechanism for modulation of expression variation. Experimental disruption of broad H3K4me3 marks altered gene expression consistency.

Evidence that genetic modulators can govern gene expression distributions suggests that evolution may act on these distributions as well as the variations in the modulating factors; however, evidence that cell-to-cell variation can be modulated does not necessarily demonstrate that the variation serves some higher-level function. In particular, robustness and consistency of cell phenotype are critical for many biological processes and organismal survival, and may provide a stronger evolutionary rational for the existence of mechanisms to control the extent of expression variation. In other words, regulatory mechanisms may have evolved under selective pressure to limit expression variation of genes that are critical to generating and maintaining cell phenotype. Regardless, the existence of these regulatory mechanisms provides avenues to experimentally manipulate cell-to-cell expression variation across cells, and so may facilitate direct testing of the above functional hypotheses.

1.6. Discussion

In this chapter, we reviewed the development of single-cell RNA sequencing measurements and discussed several theories for single-cell transcriptome heterogeneity that are prevalent in the literature (section 1.3). We then explored the idea that, in some cases, single-cell variation may be indicative of individual cells having (previously) hidden differentiated functional identities; and, that the ensemble of diverse identities is required for higher-level system function

(section 1.4). Each theory was discussed in isolation, separately from the functional hypotheses; however, further insight may be gained by considering these ideas collectively.

The prevalent theories for single-cell transcriptome heterogeneity described in section 1.3 are not mutually exclusive, and observed expression variation might be considered in the context of multiple theories. For example, expression variation due to the probabilistic nature of transcription may be apparent when redundancy in the regulatory network is present to buffer the effects of this variation (Raj et al., 2010). Additionally, no single theory will explain all observed gene expression variation. Individual cells differentiate to mature cell states, and undergo dynamic processes and homeostasis, experience contact signaling, and are locally restricted for nutrient update. Instead, it may be helpful to think about these theories in terms of several distinguishing characteristics. One such feature is the scale of reference used: the objects of focus include molecular mechanics at the level of a single gene, multi-genic characteristics of the complete transcriptome and interdependencies across genes, and the local environment and neighbors of individual cells. A second distinguishing feature is the extent of attention paid to explaining the molecular sources of expression variation and to describing the phenotypic rationale for, and effects of, this variation. This difference accounts in part the non-exclusivity of these theories. Finally, the theories differ in assumptions about the stability of the observed expression variation over time. Because single-cell RNA sequencing measurements (as well as *in situ* hybridization measurements) require the sacrifice of individual cells, analysis of expression variation using these measurements requires assumptions about the temporal dynamics of the observed expression patterns. These assumptions may have substantial effects on data interpretation. For example, by changing the assumption of expression level stability, the same observed variation might be interpreted as indicating a diversity of stable cell types or asynchrony in dynamic processes.

The population function hypotheses described above (section 1.4) might be examined in light of the theories described in section 1.3. The functional hypotheses employ an unusually broad scale of reference, considering single-cell expression variation to be a population-level characteristic with population-level phenotypic effects. In this way, these hypotheses begin to bridge the gap between single-cell characteristics and the behaviors and function of multi-cellular tissues.[2] Additionally, the validity of the above functional hypotheses, in most cases, does not depend on the sources of single-cell expression variation or their stability over time. The relationship between phenotype, at a cell- and population-level, and gene expression, however, is of particular importance to these hypotheses. At this time, methods to concurrently measure gene expression and cell phenotype are extremely rare (although see ref. Wilson et al., 2015), and defining or measuring population phenotype and function are not easy tasks. Pursuit of functional hypotheses will require progress on both fronts.

1.7. Introduction to dissertation research

The work in this dissertation lays groundwork for future studies of the role of single-cell expression variation in tissue-level function, providing an initial characterization of gene expression variation, transcriptome-wide, in five mammalian tissues and developing principled methods to study expression variation using single-cell RNA sequencing methods.

In chapter 2, we surveyed of transcriptome heterogeneity across single cells from diverse mammalian tissues using single-cell RNA sequencing data from five cell types. Because single-cell RNA sequencing is still a new technology, we rigorously evaluated control data to identify an expression level threshold beyond which our measurements are quantitative. Additionally, we

[2] Studies that focus on classification of cell types also consider this relationship, but in this case population behavior is not the object of study. These studies rely on the assumption that categorization of the stable building blocks that comprise a tissue will facilitate understanding of tissue behavior.

developed a method to estimate experimental variation as a function of gene expression level and we used this to calibrate measurements of biological variability. Implementing these methods, we found that expression variability differs across tissues, and that expression variability may be broadly conserved across rat and mouse pyramidal neurons, suggesting that patterns of gene expression across cells may be important for tissue-level function. Work in this chapter was performed in collaboration with Mugdha Khaladkar, Tae Kyung Kim, Jennifer M. Spaethling, Chantal Francis, Sangita Suresh, Stephen A. Fisher, Patrick Seale, Sheryl G Beck, Tamas Bartfai, Bernhard Kuhn, James Eberwine, and Junhyong Kim and was published in Genome Biology (Dueck et al., 2015b).

In chapter 3, we analyzed data generated as part of a large-scale collaborative control experiment to assess three different single-cell RNA sequencing methods: SmartSeq, NuGen Ovation and aRNA. We assessed measurement sensitivity, described sequence characteristics of genes that are prone to missing values, and evaluated the effect of sequencing depth on gene detection. We additionally characterized measurement precision and accuracy, overall and as a function of gene expression levels, and evaluated several protocol variants for performance improvement. This work was performed in collaboration with Rizi Ai, Ray Dominguez, Oleg Evgrafov, Jian-Bing Fan, Stephen Fisher, Chantal Francis, Jennifer Hernstein, Tai Kyung Kim, Hugo Kim, Sonia Lin, Rui Liu, Bill Mack, Neeraj Salathia, Jennifer Spaethling, Tade Souaiaia, Jai-Yoon Sul, Andre Wilberg, Robert Chow, James Eberwine, James Knowles, Kun Zhang, and Junhyoung Kim, and a manuscript describing this work is In preparation.

The dissertation concludes with Chapter 4, which contains a summary of key results and a discussion of future directions for studies of single-cell transcriptome heterogeneity.

CHAPTER 2: Cell-type specific patterns of single cell transcriptome variation

2.1. Background

The transcriptome is a key determinant of the phenotype of a cell (Kim and Eberwine, 2010) but increasing evidence suggests the possibility that large variation in transcriptome states exists across cells of the same type. High variability in single-cell transcripts have been described using various techniques, including targeted amplification (Cornelison and Wold, 1997; Miyashiro et al., 1994; Tay et al., 2010), florescent in situ hybridization or FISH (Raj et al., 2010) and whole transcriptome assays (Buckley et al., 2011; Hashimshony et al., 2012; Islam et al., 2011; Shalek et al., 2013, 2014; Tang et al., 2009). While substantial research has explored the molecular mechanisms of this variation (Kaufmann and van Oudenaarden, 2007; Munsky et al., 2012; Raser and O'Shea, 2005), a key question remains: how does this transcriptomics variation map to external phenotypic variation? Is gene expression variation explained in part by cell physiological dynamics, such as metabolic phases of the cell like circadian rhythm or cell cycle (Miller et al., 2007)? Is the expression profile of a morphologically complex neuron more variable than that of a morphologically simpler cell, such as a brown adipocyte? Is there cell-type specificity or gene-class specificity to single-cell variability? To characterize the complexity and pattern of variation at the level of single cells we carried out single-cell RNA sequencing of multiple individual cells from five different mouse tissues, as well as rat samples for two of these tissues, with high depth of coverage. Most estimates of number of mRNA molecules in a mammalian cell suggest under ~300,000 molecules per cell (Islam et al., 2011). With ~10,000 expressed genes, the average number of molecules per gene is ~30, suggesting that most of the transcriptome requires deep coverage and careful amplification for quantitative characterization. For this study, we used linear in vitro transcription for RNA amplification and quality controlled the RNA sequencing to include only those samples for which we had at least five million uniquely mapped exonic reads. Using this dataset as well as an extensive control dataset, we developed

new analytical routines to carefully characterize patterns of gene expression variability at the single-cell level and dissected the cell-type-specific variability in relation to cell identity. We find evidence that single-cell transcriptome complexity and cell-to-cell variation have cell-type-specific characteristics and that patterns of gene expression variation may be subject to regulation.

2.2. Results

2.2.1. Single-cell RNA-sequencing datasets

For each single-cell sample, we created a cDNA library after cell isolation that was linearly amplified by the antisense RNA (aRNA) method (Eberwine and Crino, 2001; Morris et al., 2011) and then sequenced on the Illumina platform. From an initial 143 cells we identified 107 high quality samples with deep genic coverage, including 13 brown adipocytes, 19 cardiomyocytes, 19 cortical pyramidal neurons and 18 hippocampal pyramidal neurons from embryonic mouse, 8 cortical pyramidal neurons and 8 hippocampal pyramidal neurons from embryonic rat, and 22 serotonergic neurons from adult mouse (Table 2.1; Table A.1). (Rat samples are included in cross-species comparisons, with primary analyses on mouse samples only. Unless otherwise specified, results are based on mouse data.) Several experimental parameters vary along with cell type, including age, collection method and culture conditions (as detailed in section 2.4.1). In fact, since individual cells are the measurement units, all of our cell-type comparisons are confounded by the natural cell-specific phenotypes, such as lipid content and cell size. Such confounding is unavoidable at this level and interpretation of transcriptome characteristics should, therefore, all include this caveat. Nevertheless, each cell-type dataset is internally consistent (Table 2.1). In the resulting dataset, the average sample has a depth of 57 million reads with 17 million uniquely aligned to exons (minimum of 5 million unique exonic reads). Using these uniquely aligned reads, we assigned read counts to RefSeq annotated genes and normalized the dataset to mitigate differences in sequencing library depth (Anders and Huber, 2010). Saturation curves generated by randomly subsampling reads for individual

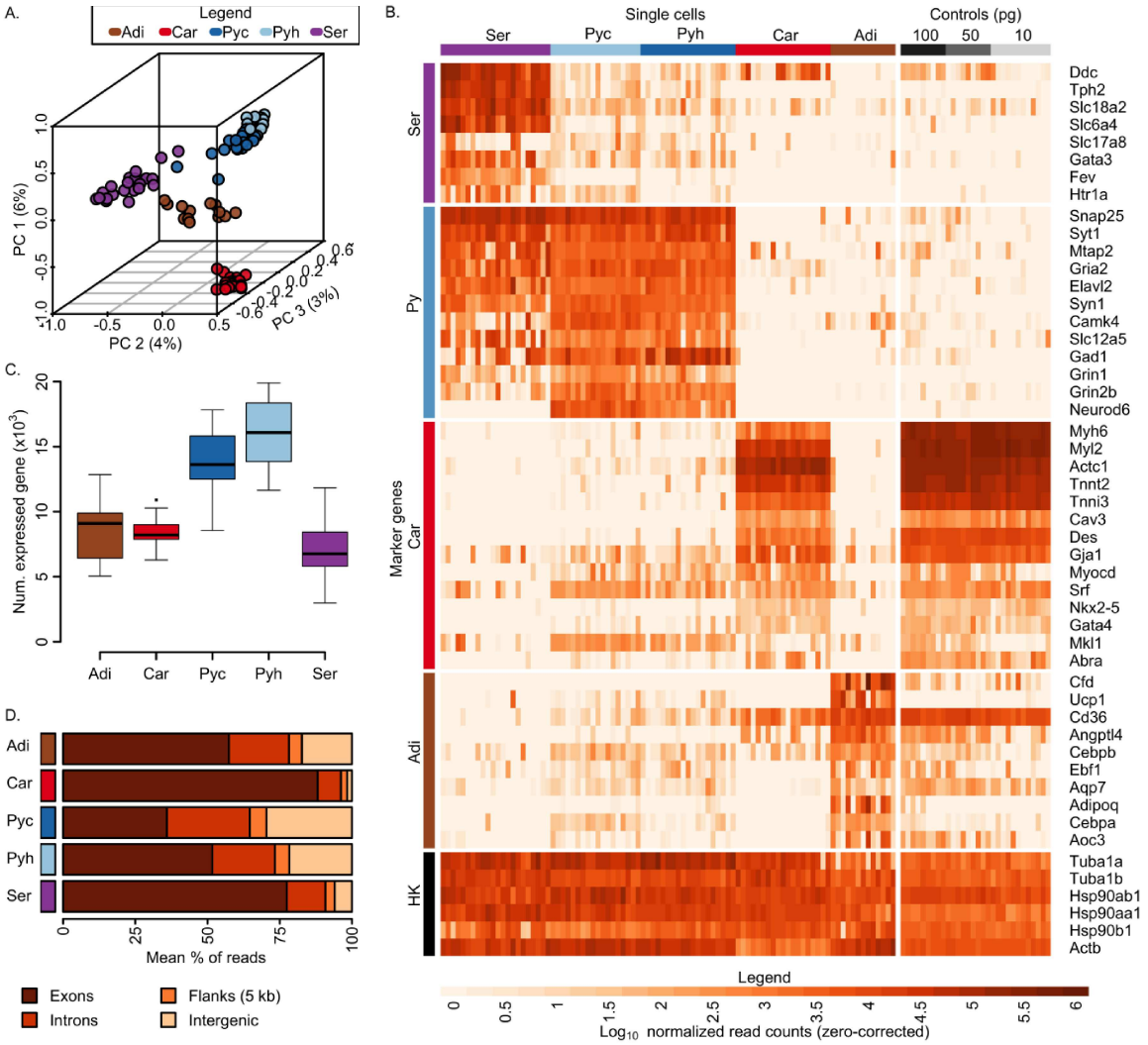
samples to generate synthetic replicates over a broad range of total depth demonstrate that little sensitivity is gained at increasing depth of coverage beyond five million unique exonic reads, suggesting sufficient depth (Figure A.1a). Additionally, within our range of coverage, we do not observe a relationship between the number of detected genes and the read depth (Figure A.1b). Principle components analysis projection of the 91 cells largely segregates the five cell types (Figure 2.1a). We examined expression levels for a curated list of marker mRNAs expected in serotonergic neurons, pyramidal neurons, brown adipocytes and cardiomyocytes to validate the quality and identity of our samples (Figure 2.1b). When clustered on these expression profiles the samples form coherent cell-type groups as expected, confirming dataset quality (Figure A.1c). While each cell type demonstrates a characteristic transcriptome profile enriched in expected marker mRNAs, we note that marker gene expression is rarely if ever limited to the expected cell type. As observed elsewhere, this suggests that multi-genic expression may better characterize cell types than expression of individual genes (Guo et al., 2010; Shapiro et al., 2013). In addition, some marker genes demonstrate substantial variability within the relevant cell type, suggesting that absolute expression levels of a small number of genes is not likely a critical determinant of cellular phenotype (Guo et al., 2010; Schulz et al., 2007). We additionally prepared 30 control samples, amplifications of bulk total cardiomyocyte RNA diluted to single-cell quantities. All dilution replicates passed quality control thresholds and the set demonstrates high pairwise correlations (Table 2.1; Table A.1 and Figure A.1d, e). Compared with single cells, the dilution controls demonstrate generally larger pairwise correlations (Figure A.1d). For details on dataset preparation, see section 2.4.4.

Table 2.1 Dataset and RNA sequencing statistics for each cell type

Sequencing statistics were based on used experimental samples. Dorsal raphe read depth based on used samples with raw data available (n=21).

Species	Cell type	Strain	Age	Cell isolation technique	Total #	# used	Ave. read depth (millions)	Ave, % aligned	Ave. % uniquely aligned	Ave. unique exonic counts (millions)
Mouse	Interscapular brown adipocyte	CD-1	E17.5	Pipette	13	13	67.5	82.2	69.0	20.8
Mouse	Cardiomyocyte	AG-Geminin	E14.5	FACS	22	19	43.4	78.8	61.9	20.0
Mouse	Cortical pyramidal neuron	C57BL/6	E18	Pipette	31	19	63.8	77.7	65.2	11.6
Mouse	Hippocampal pyramidal neuron	C57BL/6	E18	Pipette	33	18	59.0	73.1	61.7	14.5
Mouse	Dorsal raphe serotonergic neuron	e-PET-YFP	P60	Pipette	28	22	58.1	64.5	53.9	22.5
Rat	Cortical pyramidal neuron	Sprague Dawley	E18	Pipette	8	8	57.1	72.8	65.6	8.9
Rat	Hippocampal pyramidal neuron	Sprague Dawley	E18	Pipette	8	8	59.4	72.3	65.2	15.9
Mouse	Dilution series control, mouse	C57	P98	Purchased	30	30	30.8	89.4	68.9	12.8
TOTAL (Biological samples)					143	107	57.3	74.1	62.1	17.1

Figure 2.1 Single-cell dataset and transcriptome characteristics.



A. Low dimensional projection of single-cell transcriptome data. Axes were selected using principle component (*PC*) analysis of expression data. Relative frequencies of read counts were variance stabilized by arcsine transform. Genes with zero read count in all cells were excluded. Values in parentheses by each axis are percentage standard deviation explained by that axis. **B.** Expression of marker genes (rows) selected from the literature (Cahoy et al., 2008; Eguchi et al., 2008; Huang et al., 2009c; Kajimura et al., 2010; Liu et al., 2010; Pipes et al., 2006) for all mouse samples (columns). **C.** The number of expressed genes by cell type. **D.** Average percentage of reads falling within annotated exons, introns, gene flanking regions and intergenic regions by cell type. Dilution control started with 100 pg (*100*), 50 pg (*50*), or 10 pg (*10*) total cardiomyocyte

RNA. Abbreviations: *Adi* brown adipocyte; *Car* cardiomyocyte; *HK* housekeeping; *Py* pyramidal neuron; *Pyc* pyramidal neuron, cortex; *Pyh* pyramidal neuron, hippocampus; *Ser* serotonergic neuron, dorsal raphe

2.2.2. Single-cell transcriptome complexity

Averaged over all 91 cells, we observed 10,796 expressed genes per cell with 50 % of reads in a cell covering the 432 most highly expressed genes. The most abundant expressed gene comprises 2 % of reads on average, over 1000 times more than the median gene. Most expressed genes are observed in multiple cells, with a small fraction (0.027 %) of private genes expressed only in a single cell. We found pyramidal neurons (cortical and hippocampal cells) comprised a distinct transcriptome complexity group compared with the other three cell types. The number of expressed genes observed is significantly greater in pyramidal neurons (average of 14,964 genes) than in cells of the other three types (average of 7,939 genes, Welch's t-test Bonferroni-corrected $p < 0.05$; Figure 2.1c; Table A.2). Significantly more genes were covered by 50 % of reads in pyramidal neurons compared with the other cell types and pyramidal neurons had higher numbers of private genes than the other cell types (Welch's t-test Bonferroni-corrected $p < 0.05$; Table A.2; Figure A.1f, g). Pyramidal neurons as a group displayed a larger fraction of reads mapping to non-exonic regions, especially the cortical cells, for which more than 60 % of reads mapped to introns and other non-coding sequences compared with a 42 % average for all cell types (Figure 2.1d). The large portion of non-exonic sequences for all cell types is consistent with reports based on bulk data demonstrating that much of the genome is transcribed (Djebali et al., 2012). The larger percentage of non-coding sequences in pyramidal neurons is also in line with previous reports of long 3′ untranslated regions (UTRs) in the mammalian brain (Ramsköld et al., 2009), with terminals well beyond annotated ends (Miura et al., 2013), and reports of intron retention in single neurons (Buckley et al., 2011).

It is possible that difference in cell size and numbers of RNA molecules might affect detection sensitivity across cell types. While the cell sizes of the cell types used in this study have not be directly assessed, it is estimated that mammalian cells cover an eight-fold range in volume (BNID 100434 Milo et al., 2010). To adjust for this possible bias, we assumed an eight-fold difference in size between pyramidal neurons and other cell types and applied a corrected detection sensitivity threshold of eight times the minimum relative frequency observed in a given pyramidal sample, ignoring all genes below this threshold (validated on control data; Table A.2; Figure A.1h). After correction all pairwise comparisons remain significant, with the exception of the brown adipocytes and cortical pyramidal neurons pair (Table A.2; Figure A.1h).

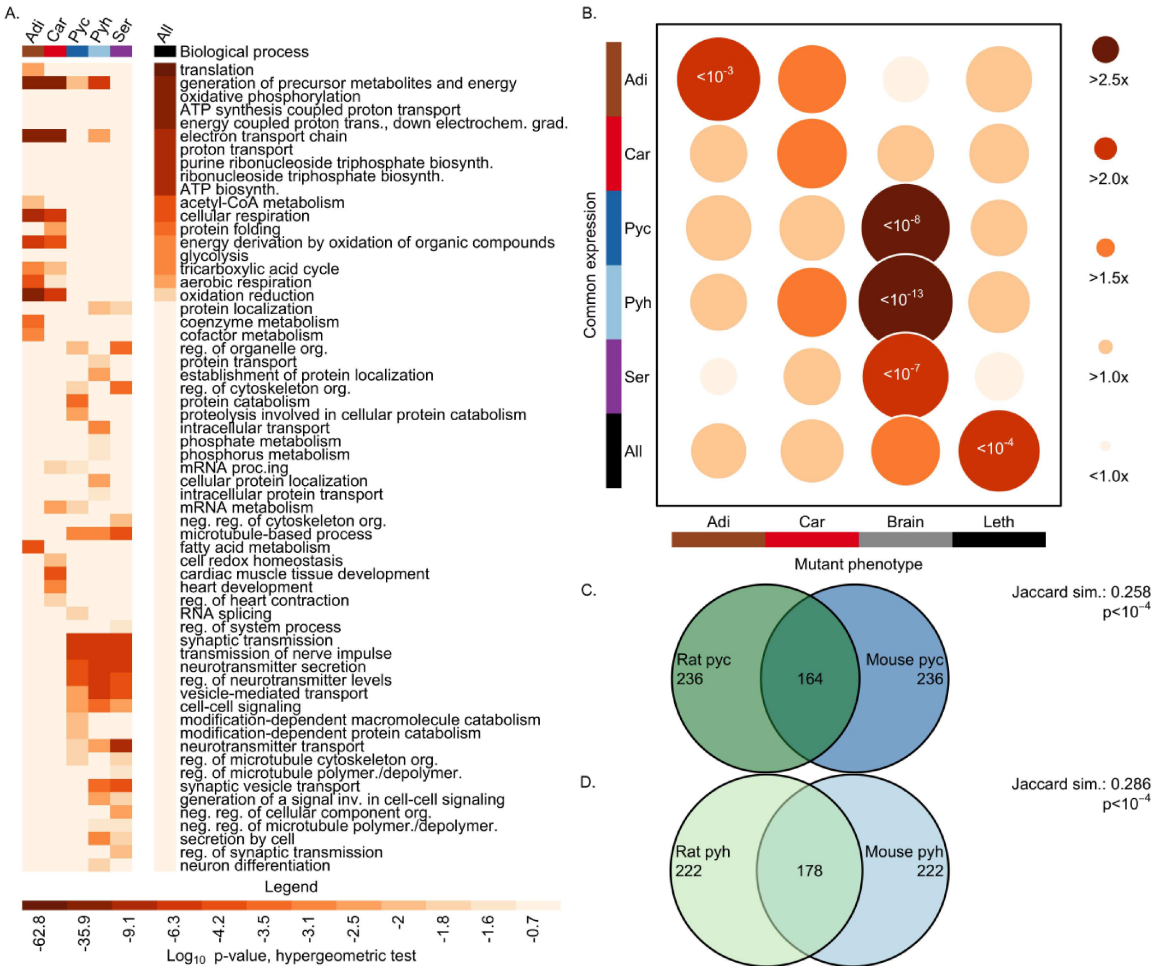
Of the 371 genes found expressed in only one cell, 334 are in pyramidal neurons. These genes include 50 olfactory receptors genes, 10 vomeronasal receptors and 9 additional genes annotated with function in cell surface receptor-linked signaling, consistent with the hypothesis that these molecules create cell diversity within the central nervous system (Table A.3; see section 2.4.5 for annotation sources). The presence of a large number of private genes, more detected genes, and greater non-coding expression all suggest unique transcriptional complexity in pyramidal neurons. While tissue studies have observed such complexity in the brain, here we identify this as a property of individual cells, not simply that of a highly diverse cellular population (Ramsköld et al., 2009). These data demonstrate that global transcriptome characteristics differ between cell types. We speculate that the broad transcription observed in single cortical and hippocampal neurons may be relevant for the phenotypic plasticity demonstrated by these cells, in contrast to the narrower functional repertoire required of heart and fat cells.

2.2.3. Consistent gene expression across single cells

Despite global transcriptional differences across cell types, we anticipated that all cells would constitutively express a subset of genes necessary for basic cell function. We identified 404 genes with evidence of expression in all 91 single cells (Table A.3). Indeed, this set is

enriched in functional annotations associated with housekeeping genes (hypergeometric test Bonferroni-corrected $p < 0.05$; Figure 2.2a). We reasoned that if this gene set is critical for basic cell function, then gene disruption should be highly detrimental at an organismal level. We found that genes commonly expressed across all cell types are significantly more likely than remaining genes to be categorized as prenatal lethal, consistent with previous suggestions that genes whose deletion is lethal demonstrate low expression noise (Figure 2.2b; Table A.4) (Kaufmann and van Oudenaarden, 2007).

Figure 2.2 Phenotypic importance of genes with consistent expression.



A. Enriched Gene Ontology biological process categories across commonly expressed genes. Category abbreviations: *biosynth*. biosynthetic; *cmpd*. compound; *depolymer*. depolymerization; *deriv*. derivation; *dev*. development; *gen*. generation; *metab*. metabolites; *org*. organization; *oxid*. oxidation; *polymer*. polymerization; *prec*. precursor; *proc*. process; *reg*. regulation; *synth*. synthesis. **B.** Association of common gene expression (rows) and mutant phenotypes (columns). Mammalian Phenotype Ontology phenotypes were grouped for affecting brown adipose tissue (*Adi*), myocardial tissue (*Car*), or brain tissue (*Brain*), or causing prenatal lethality (Table A.7) (Eppig et al., 2012; Smith and Eppig, 2009). Circle size indicates enrichment of phenotypic category in commonly expressed genes. Bonferonni-corrected p values are included for significant chi-square tests. **C, D.** Overlap of common genes across species for cortical (C) and hippocampal (D) pyramidal neurons. P values were calculated by random sampling (see section 2.4.6). Sample sizes and abbreviations: brown adipocyte (n = 13, *Adi*); cardiomyocyte (n = 19, *Car*); pyramidal neuron, cortex (mouse n = 19, rat n = 8, *Pyc*); pyramidal neuron, hippocampus (mouse n = 18, rat n = 8, *Pyh*); serotonergic neuron, dorsal raphe (n = 22, *Ser*)

To examine commonly expressed genes within each of the five cell types, we selected the 400 highest expressed genes (defined by the minimum value in any cell) for each cell type, excluding the 404 universally expressed genes (Table A.3). Because these gene subsets demonstrate common expression within cells of each type, we hypothesized phenotypic importance. As expected, this set of genes is enriched in annotations associated with cell-type-specific function, such as fatty acid metabolism, cardiac muscle tissue development, neuron differentiation and synaptic transmission (hypergeometric test Bonferroni-corrected $p < 0.05$; Figure 2.2a). These genes are significantly more likely than expressed background to produce tissue-specific phenotypes on mutation but not to result in prenatal lethality (Figure 2.2b; Table A.4), with the exception of cardiomyocytes, which are proliferating and whose highly expressed genes are dominated by cell cycle function. The majority of commonly expressed genes, within each cell type and those that are universal in expression, do not have published phenotypic annotation and present a potential resource for disease association studies.

If common expression across single cells is indicative of critical gene function, this expression pattern may be conserved across species. We identified commonly expressed genes in rat cortical and hippocampal pyramidal neurons and compared them with the commonly expressed genes in mouse cortical and hippocampal pyramidal neurons (restricting analysis to unambiguous homologues and excluding homologues of universally expressed genes). The identities of cell-to-cell commonly expressed genes in each species show highly significant overlap (random sampling $p < 10^{-4}$; Figure 2.2c, d). That is, if a mouse gene tends to be commonly expressed in all pyramidal neurons (but not in every cell type), its rat homolog also tends to be commonly expressed, providing additional support that commonly expressed genes perform critical functions.

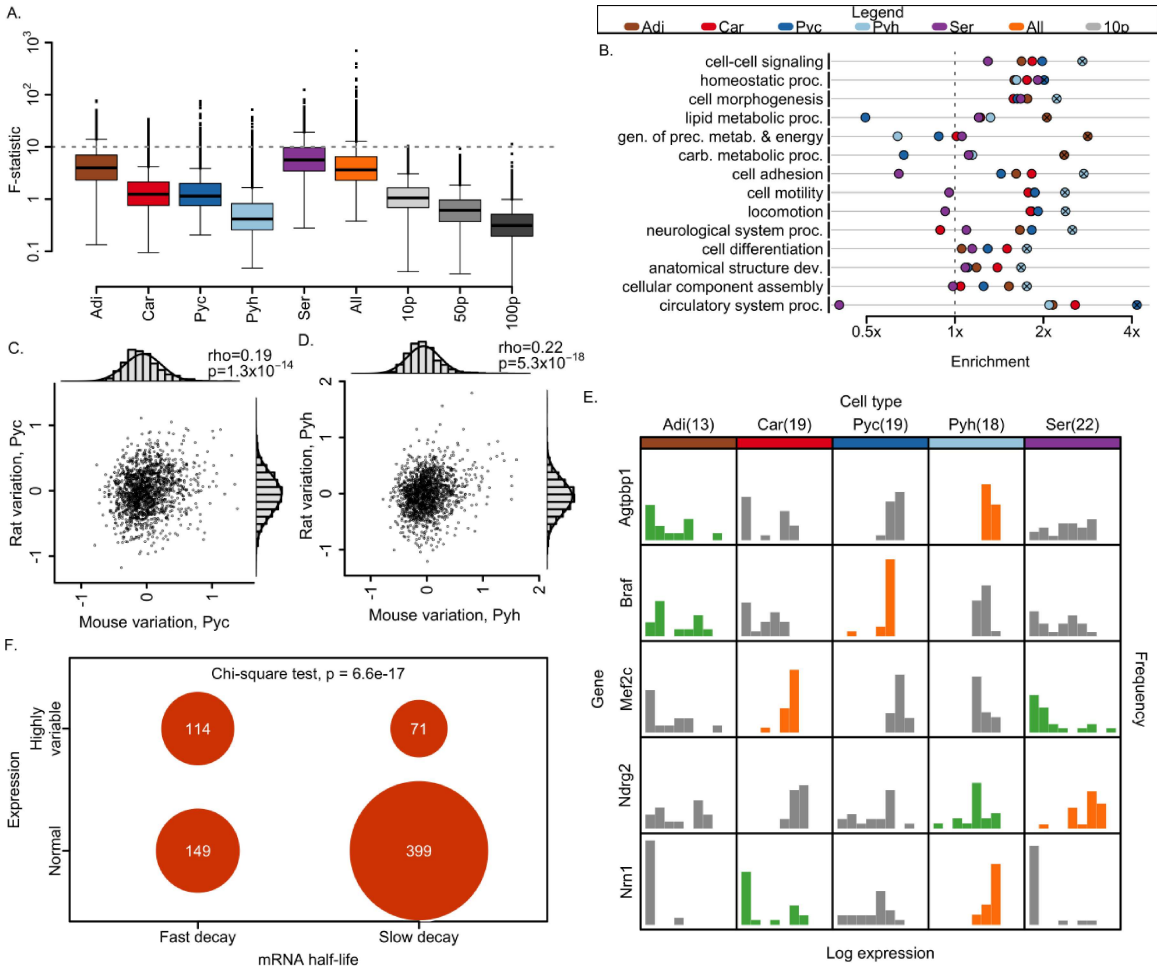
2.2.4. Single-cell transcriptome variation

Because technical variability from single-cell RNA-sequencing measurements depends on their expression levels (Brennecke et al., 2013), we used the dilution replicates to determine a reliable range of gene expression before performing quantitative analyses (Figure A.2a–d). Examining replicates beginning with 10 picograms (pg) of total RNA (comparable to a single cell) (Ramsköld et al., 2012), we identified an expression level that meets four reliability criteria: (1) at least 50 % of genes have no missing values in all dilution replicates (Figure A.2a); (2) variation across all single-cell samples is larger than twice that observed across dilution replicates (Figure A.2b); (3) variation across dilution replicates is approximately normally distributed (Figure A.2c); and (4) log read depth to log expression rank shows a consistent functional relationship (Figure A.2d). All criteria result in similar expression level thresholds: relative frequencies ranging from 2.6×10^{-5} to 6.3×10^{-5}, corresponding to 328 to 789 reads. These conservative thresholds indicate reliable quantification of around four to nine input molecules. Approximately 25–37 % of the expressed transcriptome in a single cell is expected to have more than four to nine molecules (assuming 150,000 total mRNA molecules in a cell). For analyses below, we excluded genes with expression level below the most stringent threshold (relative frequency of 6.3×10^{-5}). Note that

use of thresholds based on replicates beginning with 10 pg of total RNA is conservative: thresholds selected for dilution replicates beginning with 50 or 100 pg of total RNA occur at lower expression levels (Figure A.2a–d).

To characterize the extent of single-cell expression variation within each cell type, we calculated the ratio of biological variation over experimental variation observed at matched expression level (studentized F-statistic). This metric is a measure of total variation across single-cell samples (a combination of biological and experimental variation) relative to experimental variation. At a given expression level, larger values indicate larger biological variation. Briefly, we summarized the dependence of experimental variation (calculated across 10 pg dilution replicates) on expression level by computing a sliding window median (Figure A.2e). We scaled variation across single-cell samples by this sliding median value at matched expression level (Figure A.3a–f). As a negative control, we also calculated this statistic for dilution controls (Figure A.3g–i). For further details, see section 2.4.7. We note that experimental variation (for the dilution controls, a combination of dilution and technical variation) depends on the gene-specific levels of RNA and the total amount of RNA (Figure A.3g–i; Figure 2.3a). Differences in cell size could confound F-statistic distribution with total RNA effects. Because differences in total RNA molecules numbers for the different cell types are unknown, we use the F-statistic only to examine relative differences in gene expression variation within each cell type.

Figure 2.3 Cell-type patterns of transcriptome variability



A. Distribution of expression variability across the transcriptome by cell type. **B.** Enrichment of Gene Ontology categories among variable genes by cell type (Ashburner et al., 2000; Carbon et al., 2009). *Crosses* indicate significance (Fishers exact test Bonferroni $p < 0.05$). Category abbreviations: *carb*. carbohydrate; *dev*. development; *gen*. generation; *metab*. metabolites; *prec*. precursor; *proc*. process. **C, D.** Partial correlation of F-statistic across species, controlling for gene expression level, for cortical (C) and hippocampal (D) pyramidal neurons. Axes are a measure of variation, controlled for gene expression level (see section 2.4.7 for details). *rho* indicates the partial correlation coefficient. *P* values are from a two-sided t-test of association. Marginal histograms are shown overlaid with a normal curve. **E.** Distribution of expression values by cell type for selected genes demonstrating highly variable expression by the outlier-sum statistic in one cell type and as following a normal distribution across cells in another cell type.

Histograms of genes identified as highly variable in a given cell type are colored *green*; those of genes identified as normally expressed in a given cell type are colored *orange*. **F.** Contingency table of gene categorization as hypervariable and as fast decaying (Schwanhäusser et al., 2011). Sample sizes and abbreviations: brown adipocyte (n = 13, *Adi*); cardiomyocyte (n = 19, *Car*); pyramidal neuron, cortex (mouse n = 19, rat n = 8, *Pyc*); pyramidal neuron, hippocampus (mouse n = 18, rat n = 8, *Pyh*); serotonergic neuron, dorsal raphe (n = 22, *Ser*); all single cells (n = 91, *All*); 10 pg dilution controls (n = 12, *10p*); 50 pg dilution controls (n = 9, *50p*); 100 pg dilution controls (n = 9, *100p*)

2.2.5. Within-cell-type variability

With the exception of hippocampal pyramidal neurons, all cell types demonstrate significantly greater transcriptome variability than that observed across 10 pg dilution controls (Wilcoxon rank-sum test $p < 0.05$; Figure 2.3a). Every cell type contains highly variable genes with an F-statistic greater than 10 (Figure 2.3a), indicating the presence of highly variable genes for each cell type. To compare the extent of transcriptome variation across cells of the same type with transcriptome differences across cell types, we computed the variance within each cell type as well as the variance between each pair of cell types (Table A.5). We found that expression variation within some of the cell types is comparable to that observed between some of the cell types: within cells of the same type, variance ranges from 0.21 to 1.31, while between cell types variance ranges from 0.26 to 1.88. This result may be affected by the difference in total RNA content of each cell. Nevertheless, this suggests that a great diversity of transcriptome states may support an equivalent external cell phenotype.

To assess whether variability might be related to cell-type-specific sub-states, we identified the 5 % most variable genes in each cell type by the F-statistic (Table A.3 and Figure A.3a–i). We tested these genes for enrichment of Gene Ontology molecular function and biological process categories (Figure 2.3b; Table A.6). Functional categories relevant for plastic phenotypes are enriched among variable genes in a cell-type-specific manner. In pyramidal neurons, variable genes are enriched in functions important to cell migration, such as cell

morphogenesis and locomotion. Generation of precursor metabolites and energy is enriched among variable genes in brown adipocytes. But we note similar enrichment is also seen among 50 pg and 100 pg dilution controls, which may be due to the effect of expression level on the F-statistic (Table A.6). This suggests that, for a subset of genes, the observed variability is due to cell-type-specific molecular physiology.

We also examined the degree of expression variation among different functional classes of genes by calculating the F-statistic for several broad categories (Table 2.2). Genes categorized with classic housekeeping functions (mitochondrial or ribosomal function) demonstrate low expression variability, while transcription factors, important in responding to the environment or modulating cell function over time, demonstrate significantly greater gene expression variability. Interestingly, ion channels demonstrate the largest variability, significantly larger than all other examined categories, possibly suggesting homeostatic modulation (Marder, 2011; Marder and Goaillard, 2006).

Table 2.2 Gene expression variability of different functional gene categories

Gene category 1	Gene category 2	Difference in adjusted mean expression variability (log_{10} F-statistic)		
		Category 1 – category 2	95 % CI	*p* value
Ion channel	Metabolism	0.32	0.22–0.41	$<10^{-5}$
Ion channel	Ribosome	0.33	0.23–0.42	$<10^{-5}$
Ion channel	Transcription factor	0.16	0.07–0.25	4.84×10^{-5}
Transcription factor	Metabolism	0.16	0.11–0.20	$<10^{-5}$
Transcription factor	Ribosome	0.17	0.12–0.22	$<10^{-5}$
Ribosome	Metabolism	−0.01	−0.06–0.04	0.92

Comparison of gene expression variability across functional gene categories controlling for gene expression level. Adjusted mean values were calculated using a two-factor ANCOVA of log_{10} F-statistic, with functional gene category and cell type as independent factors and conditioning on gene expression level (log_{10} average normalized read depth). Adjusted means are reported at the

average gene expression level. Reported p values are for a two-sided Tukey's test and are calculated based on a joint t-distribution to control the family-wise error rate.

If gene expression variation amongst individual cells is important for tissue function, the degree of variation itself may be conserved across species. We calculated the F-statistic for cortical and hippocampal pyramidal neurons in rat, filtering genes by the quality control threshold described above. For each cell type we computed the partial correlation of the F-statistic across species, controlling for gene expression levels to ensure that correlation was not simply due to shared levels of gene expression (see section 2.4.7 for details). The partial correlation coefficient across species is significant for both cell types examined (two-sided t-test of association $p < 10^{-13}$; Figure 2.3c–d). The number of cells from rat is relatively small and further studies are required to confirm that gene expression variation is conserved across species. Furthermore, additional data are needed to dissect whether such conservation in variance is cell-type-specific or indicative of more general selection for tight regulation of those genes that are critical for global cell function. Lastly, as in our other results, any statement on variances must be interpreted cautiously because of the intensity-dependence of variance and analytic techniques for variance stabilization. But the current comparative data are consistent with the hypotheses that gene expression variation is regulated, at least for some genes, and that the pattern of gene expression across a population of cells might be important for tissue function.

2.2.6. Patterns of expression variation

We next examined a subset of genes with patterns of extreme variability across cells of the same type. To identify these genes, we used the outlier-sum statistic, a method proposed to detect genes with high expression in a subset of samples (Table A.3) (Hellwig et al., 2010; Tibshirani and Hastie, 2007). As controls, we also identified variable genes across all cell types (a positive control) and across dilution replicates (a negative control) by the same method. Genes identified across all cell types are enriched for categories relative to cell-type-specific functions,

including heart development, behavior, and regulation of system process, a category encompassing processes that modulate tissue function (Table A.6). No functional categories are enriched among genes identified across dilution replicates. Genes identified across cardiomyocytes, which were collected from embryonic tissue undergoing cell division, are enriched for function in mitosis, nuclear division and organelle fission; and genes identified in brown adipocytes, which generate heat and fever and share some immune regulators, are enriched in immune response related genes. While only a small number of genes were identified in pyramidal neurons, limiting the ability to detect functional enrichment, individual genes demonstrate tissue-specific function. The most highly ranked genes in cortical neurons include Crh, with function in associative learning, long-term memory and response to cocaine; Vip, again with function in learning and memory; and Tac2, involved in associative learning and long-term memory formation in humans (Table A.3). The functional coherence among these genes again raises the possibility that single-cell variation, at least in a subset of genes, is regulated.

We identified 58 genes demonstrating qualitatively different expression patterns in different cell types. Each of these genes is classified as highly variable by the outlier-sum statistic in at least one cell type, and as following a normal distribution across cells in at least one other cell type Figure 2.3e; Table A.3). Each gene is highly expressed in multiple cell types (due to use of quality control threshold), within the top 36 % of expressed genes, but demonstrates a markedly different expression distribution in different cell types. This difference of expression pattern of the same gene in different cell types also suggests that this expression variation may be controllable and the result of regulation.

For this set of 58 genes we do not find significant enrichment for any Gene Ontology classification. But we observed individual cases of genes of note that are associated with cell-type-specific function. For example, mutants of *Nrn1*, a gene following a normal distribution across hippocampal neurons, are associated with abnormal spatial learning and impaired

contextual conditioning behavior. Mutations of *Braf*, a gene following a normal distribution across pyramidal neurons, are associated with abnormal learning and abnormal hippocampal granular and cerebral cortex pyramidal neuron morphology. *Mef2c* follows a normal distribution across cardiomyocytes; this gene functions in cardiac muscle development and has been used in transdifferentiation experiments converting fibroblasts to cardiomyocytes (Vierbuchen and Wernig, 2012). These examples again suggest that consistency in gene expression may be an indicator of critical phenotypic relevance.

2.2.7. Association of extreme variability with RNA half-life

Because variation in gene expression may be buffered by long RNA half-life, we hypothesized that if extreme variability is generated by transcriptional switching, these genes may demonstrate rapid decay. We categorized genes as having slow or rapid decay based on publically available RNA half-life measurements and then tested for an association with variability (Schwanhäusser et al., 2011). Genes identified as variable are significantly more likely to be classified as rapidly decaying than genes with highly consistent expression (Chi-square test $p < 10^{-16}$; Figure 2.3f). While a rapid decay does not necessarily indicate large expression variability, highly variable genes rarely have slow decay and genes with slow decay are rarely highly variable (Figure A.3j). This is consistent with models of transcription, which suggest that bursts of changes in RNA numbers may occur when the intervals of inactive transcription are long relative to mRNA decay (Munsky et al., 2012). A rapid decay rate of a transcript may enable rapid changes in gene expression levels.

2.3. Discussion

Cell-type-specific characteristics of the single-cell transcriptome recapitulate characteristics of tissue-level expression data, indicating that, at a minimum, single-cell RNA sequencing is meaningful at the level of cell-type identity. In addition, our data suggest that expression variation among individual cells of the same type has biological significance. Figure

2.1a, a principle components analysis projection of the 91 mouse cells, displays the overall pattern of transcriptome variation and suggests a complex pattern of single-cell variability both within each of the five cell types and between them. The analyses described above suggest that these expression patterns may be driven by a multitude of factors. The enrichment of certain functional classes among variable genes suggests that some observed variability is due to measuring cells in different phases of functional dynamics. The greater transcriptome complexity we see in pyramidal neurons, both in the number of expressed genes and in the degree of non-coding expression, suggests that the complexity of available RNA may be related to the morphological complexity and plasticity of a cell. Enrichment of cell surface receptor molecules among private genes suggests that some variation at the level of individual cells might be "programmed" to induce a heterogeneous assembly of cells in a tissue, perhaps to carry out organ level function.

Single-cell transcriptome measurements are becoming increasingly common with a variety of techniques for RNA amplification and sequencing. In particular, we emphasize the need for careful control data as well as statistical analysis routines that incorporate the unique properties of single-cell transcriptomes. Another key aspect of inference from single-cell data is the depth of read coverage required to quantitatively characterize most of the transcriptome. Using statistical analysis of the control data and other characteristics of our data we estimate that with the aRNA technique we can conservatively measure with quantitative precision four to nine molecules of input mRNA, which is approximately 25–37 % of the genes that are expressed in a single cell. A difficult problem is that measurement variation is likely a function of absolute numbers of molecules. Individual cell characteristics, such as the number of RNA molecules, size of the cell, and cell components that interact with RNA recovery, will all confound technical variability. Any given measure will include interactions of both biological and technical factors. This is a problem that will be endemic to any single-cell quantitative measurements.

Summarizing our data, the functional coherence of genes identified based on expression variation, tissue-level phenotypes in animals with mutational knock-outs of consistently expressed genes, and correlation of expression patterns across rat and mouse leads us to hypothesize that some observed variation is necessary for tissue-level function of the cell and that the degree of single-cell variation in gene expression may be under regulatory control. In addition, we hypothesize that the same external cell phenotype may be robustly produced from a great diversity of transcriptome states, as long as these states remain within some bounds. Gene expression is required to maintain the multi-genic stoichiometric constraints of a cell's normal physiology but such constraints may allow many degrees of multi-dimensional freedom (Rifkin et al., 2000). Thus, we propose there are both functional and degrees-of-freedom rationales for the high degree of single-cell transcriptome variation and we suggest that cells within organs may be more like individuals in an ecological community rather than homogeneous replicate units.

2.4. Materials and methods

2.4.1. Cell culture and single cell isolation

Primary cultures of embryonic day 18 (E18) hippocampal and cortical neurons from mouse (C57BL/6, Charles River Laboratories, Inc.) and rat (Sprague Dawley) were cultured as previously described (Buchhalter and Dichter, 1991). Interscapular brown adipose tissue was extracted from E17.5 CD-1 mice and cultured for one day as described elsewhere (Spaethling et al., 2015). We identified pyramidal neurons and brown adipocytes by cell morphology, isolated single cells with patch pipettes, deposited collected material into an eppendorf tube and froze it immediately at −20 C° until storage at −80 C° (Morris et al., 2011). Serotonergic neurons, identified with yellow fluorescent protein (YFP) expression, were isolated by pipette directly from acute slices of the dorsal raphe of P60 ePet-YFP mice as described elsewhere (Scott et al., 2005; Spaethling et al., 2015). Ventricular cardiomyocytes were isolated from E14.5 transgenic mice expressing the green S/G2/M fluorescent ubiquitination-based cell cycle indicator "Fucci"

(Sakaue-Sawano et al., 2008) using a modified protocol of the neomyts cardiomyocyte isolation kit (Cellutron Life Technologies). Cycling cardiomyocytes were selected by flow cytometery based on expression of the Fucci marker and mAG-*hGem* (1/110) transgene. Cells were sorted into 96-well plates with reverse transcription buffer for linear amplification as described below. For full details on cardiomyocyte selection, see (Dueck et al.).

2.4.2. mRNA amplification and library construction

Collected samples were individually amplified using three rounds of a linear in vitro transcription-based method described elsewhere (Morris et al., 2011). Amplified material was quantified and size-checked using a Bioanalyzer RNA Nanochip (Agilent), then prepared for multiplexed paired-end sequencing using the TruSeq or mRNA-Seq systems according to the manufacturers' instructions. Initial mRNA selection steps were skipped to accommodate aRNA amplified material. Samples were sequenced on HiSeq instruments to produce 100-base paired-end reads. Sample-specific sequencing data can be found in Table A.1.

2.4.3. Dilution controls

To control for technical variation arising during amplification and sequencing preparation, we performed replicate amplifications of bulk RNA diluted to near single-cell quantities. The starting bulk sample was heart tissue total RNA from a C57/BL6 adult male mouse (Zyagen). Twelve amplifications were begun with 10 pg of total RNA, nine with 50 pg and nine with 100 pg. After three rounds of aRNA amplification quality and quantity of all samples were assessed using Nanodrop and Bioanalyzer RNA Nanochip. Samples were then prepared for sequencing using the stranded TruSeq mRNA protocol (Illumina). Replicates were sequenced as above.

2.4.4. Alignment, quantification and sample selection

For mouse samples, we trimmed reads for adapter and poly(A) contamination using in-house software before aligning to the mouse genome and transcriptome using RNA-Seq Unified

Mapper (versions 2.0.2_06 and 2.0.3_04) and mouse genome build mm9 (Grant et al., 2011). Uniquely aligning reads with three or fewer mismatches per 100 bases were retained for further analysis. Using RefSeq annotations downloaded from the UCSC genome browser in December, 2012, we assigned read counts to genes using HTSeq and htseq-count (Anders et al., 2015; Karolchik et al., 2004). Only exonic reads were counted, with overlap assigned using the intersection non-empty method. Note that the use of uniquely aligned exonic reads means that genes with limited unique sequence will be largely missed in this analysis. Though this is rare, one notable example is the housekeeping gene Gapdh. Rat samples were aligned with STAR (Dobin et al., 2012) to rat genome build rn5. Processing after alignment was completed as described above with rn5 RefSeq annotations for gene quantification (downloaded from the UCSC genome browser).

We further filtered samples to retain those with high genic read coverage (greater than five million uniquely mapping exonic reads) and to create an overall dataset with similar sequencing library characteristics. Every included mouse sample was required to have at least two out of the three following traits: at least 50 % uniquely mapping reads, at least 1500 genes covered at reasonable depth (greater than ten reads), and less than 80 % short fragments (inferred from the percentage of uniquely mapping read pairs where the mate pair alignments overlap). Because the statistic used to identify short fragments is distinctive to RUM alignment reports, we report no comparable statistic for the STAR-aligned rat samples. For these samples, we required at least one of the two remaining criteria. Quality information for each sample can be seen in Table A.1.

To mitigate differences in read counts due to variable sequencing depth, we normalized the resulting dataset by the method proposed by Anders and Huber (Anders and Huber, 2010). In this method, a pseudo-reference sample is generated by taking the geometric mean across samples of ubiquitously expressed genes. A size factor estimating the contribution of sequencing

depth is estimated as the median ratio of expression of ubiquitously expressed genes in a single sample to the pseudo-reference. All read counts for that sample are then adjusted by this factor. All further analyses, except those using relative frequencies, used normalized read counts.

Because both GC content and gene length have been shown to affect RNA-sequencing measurements, we tested for a relationship between these gene traits and normalized gene counts (Figure A.2f) (Zheng et al., 2011). We tabulated gene length and GC content for mouse RefSeq genes, with the exception of 43 genes with annotated positions on multiple chromosomes. Correlations of these traits with gene counts are negligible.

We performed principle component analysis of gene expression data after variance stabilizing relative frequencies by arcsine transform. Genes with zero read count in all cells were excluded. To visualize a projection of the data on the three components with largest singular values, we used the R 'scatterplot3d' library (Ligges et al., 2014).

2.4.5. Characterization of single-cell transcriptomes

All statistics were computed in R (R Development Core Team, 2010). t-Tests used to test the null hypothesis of no difference in means were performed for groups with different sample sizes and different variance, with rejection of the null hypothesis at a Bonferroni-corrected p value of 0.05. Expressed genes include any gene with at least one uniquely aligned read. Private genes include any gene with at least one uniquely aligned read in a single cell, but none in any other cell. The functional and phenotypic annotations for gene sets throughout the paper were found via Mouse Genome Informatics or DAVID Functional Annotation Tool, using Gene Ontology molecular function and biological process classification, as well as Mammalian Phenotype Ontology classification (Ashburner et al., 2000; Eppig et al., 2012; Huang et al., 2009a, 2009b; Smith and Eppig, 2009). Mammalian Phenotype Ontology annotations were accessed through Mouse Genome Informatics on June 2013, excluding annotations based exclusively on cell line

experiments (Eppig et al., 2012; Smith and Eppig, 2009). Gene Ontology annotations for all expressed genes were also downloaded June 2013 from Amigo (Carbon et al., 2009).

To characterize cell-type patterns of transcription genome-wide, we first classified regions of the autosomal genome as exonic, intronic, flanking or intergenic. To do this, we accessed the following annotations for the mm9 reference genome from the UCSC genome browser in March 2013: miRBase, RefSeq genes, Ensembl Genes, Vega genes and UCSC known genes (Karolchik et al., 2004; Kent et al., 2002). Any region annotated as an exon for a gene or non-coding RNA from any of these sources was classified as exonic. Regions internal to transcribed units but not annotated as exonic by any annotation were classified as intronic. Regions 5 kilobase pairs upstream and downstream of any transcribed unit, or up to the nearest neighboring exonic region, were categorized as flanking regions. All remaining genomic regions were categorized as intergenic. Reads were assigned to these regions using HTSeq via the intersection non-empty method as above. For each sample, the fraction of reads assigned to each region was calculated and the mean value for each cell type is shown.

2.4.6. Consistent genes

Genes with universal expression were defined as genes with at least a single read in all samples. Enriched functional terms were found using DAVID Functional Annotation Tool to be significant relative to *Mus musculus* background by hypergeometric test at a Bonferroni-corrected p value of 0.05, using Gene Ontology biological process FAT annotations (Ashburner et al., 2000; Huang et al., 2009a, 2009b). In cases where the identical gene subset was enriched in multiple terms, the most significantly enriched term(s) was (were) reported. As for universally expressed genes, commonly expressed genes were found for each cell type. Of all commonly expressed genes within a particular cell type, the top 400 genes were selected by the minimum read count in any cell excluding universally expressed genes. Functional enrichment for these gene sets was found as above. There were 3752 expressed genes with Mammalian Phenotype Ontology

annotations and these were used to test the association between common expression and prenatal lethality and between common expression and tissue-specific mutant phenotypes. For annotations assigned to these phenotypic categories, see Table A.7. Because a very small number of expressed genes had phenotypic annotations exclusive to a single brain tissue, we excluded brain tissue-specific terms and instead focused on terms with broader neuron or brain phenotypes. Association was tested using chi-square tests, rejecting the null hypothesis of no association at a Bonferroni-corrected p value of 0.05. Enrichment was calculated as the fraction of common genes with phenotype relative to the same fraction for remaining genes.

To compare common genes across species, we downloaded homologue annotations from Mouse Genome Informatics in February 2015 (Eppig et al., 2012; Smith and Eppig, 2009) and filtered homologues to include only unambiguous cases with one assigned gene in each species. Common genes were identified in each cell type as described above separately for each species, excluding homologues of mouse universally expressed genes. For each cell type, we calculated the Jaccard index of identified common genes as a measure of gene list similarity. To determine whether the observed similarity was significant we randomly sampled gene lists of matched size from each species, computing the Jaccard index for each. The assigned p value is the fraction of observed similarities of the same or greater value than the true index in 10,000 random samples.

2.4.7. Gene expression variability

Our aim was to identify biological variation across single cells and limit the effect of technical variation on our conclusions. Because single-cell RNA-sequencing experiments are subject to technical variation dependent on expression level (Brennecke et al., 2013; Ramsköld et al., 2012), we generated an estimate of experimental variation as a function of expression level to use as a baseline measure of technical variation. We generated this control curve using a kernel approach, first computing variation across twelve 10-pg dilution replicates for all genes, then

summarizing the variation at a given expression level as the median value across 500 neighboring genes (Figure A.2e). We then used an F-statistic as our measure of variation across single-cell samples, scaling observed variation by control variation at matched expression level. Specifically, for gene gi with average expression level xi, we calculate:

$$F_i = V_{total}(g_i, x_i) / V_{exp}(x_i)$$

where V_{total} is the total sample variance calculated on relative frequencies for a given cell type, and V_{exp} is an estimate of experimental variation as a function of expression level, as described above. Because the total observed variance is a combination of biological and technical variation, larger values for this measure at a given expression level indicate larger biological variation. As an additional negative control, we also computed the F-statistic for the dilution replicates against the control curve and include this group in variation analyses.

The F-statistic is sensitive to non-normality, but it is not generally true that experimental variation in single-cell RNA sequencing is normal, particularly for genes observed only in a subset of replicates. We expected that there may be an expression level beyond which missing data rarely occurs and experimental variation is approximately normal. To identify such a threshold, we computed the Shapiro-Wilk statistic separately for all genes observed in 10-pg dilution replicates. As for the variation control curve, we found the median curve for Shapiro-Wilk p values and identified the expression level beyond which all median p values are greater than 0.01 across dilution replicates (Figure A.2g). The selected threshold (relative frequency of 6.32×10^{-5}) corresponds approximately to a read depth threshold of greater than 789 reads, resulting in retention of less than 25 % of genes for each cell type. Genes with average relative frequencies below this threshold for a given cell type were excluded from variation analysis. We additionally identified expression level thresholds that satisfy three further quality-control criteria (see section 2.2.4; Figure A.2a–d). All produce similar, though slightly lower, expression level thresholds. We excluded genes with mean expression below the most stringent threshold from all variation

analysis. We note that thresholds identified based on dilution controls beginning with larger amounts of input RNA occur at lower expression levels and are less stringent than the threshold used.

As for gene counts, described above, because both GC content and gene length have been shown to affect RNA-sequencing measurements, we checked for a relationship between these gene traits and the above-described F-statistic and found that correlations between the traits and measure are negligible (Figure A.2h). We additionally examined the relationship of the F-statistic with expression level (Figure A.3a–i). The measure is not strongly dependent on expression level, though for biological samples the largest F-statistic values occur at the highest gene expression levels. This is appropriate, since biological signal will be most clearly distinguishable from experimental variation at high expression levels; however, genes ranking within the top 5 % by F-statistic value above the expression level threshold are generally found to span a broad range of expression levels (Figure A.3a–f).

To test whether functional gene categories demonstrate different degrees of expression variability, we compared mean F-statistic values across genes categorized as metabolic, ribosomal, transcription factor, or ion channel. Annotations for ion channels were accessed from the International Union of Basic and Clinical Pharmacology public database (downloaded June 2013) (Sharman et al., 2012), and for transcription factor activity, ribosomal function, or metabolic function from Gene Ontology. Because the F-statistic depends on expression level (see above) and because different functional gene categories may have different mean expression levels, we compared expression variability across categories while controlling for gene expression level using a two-factor ANCOVA (see Table 2.2). Several R packages (car, effects and multcomp) were used to perform this analysis (Fox, 2003; Fox and Weisberg, 2011; Hothorn et al., 2008).

To compare expression variability across species, we filtered genes to retain those with unambiguous (single) homologues in both rat and mouse, as described above. Additionally, only

genes passing the quality control expression level threshold in both species were considered. Because the F-statistic depends on expression level (see above), we wished to control for expression level in measuring expression variability similarity because we anticipate that expression level may be conserved. For this reason, we used partial correlation as a measure of similarity across species, conditioning on the mean expression level for each species. Specifically, for each species we fit the model:

$$\log_{10}(\text{F-statistic}) = b_0 + b_1\log_{10}(X_{mouse}) + b_2\log_{10}(X_{rat})$$

where X_{mouse} and X_{rat} refer to mean expression values. Partial correlations were calculated using residuals. Significance was calculated using a two-sided t-test for association, rejecting the null hypothesis of no association at $p < 0.05$.

To identify genes with patterns of extreme variation within each cell type, we employed the outlier-sum statistic (Hellwig et al., 2010; Tibshirani and Hastie, 2007). Briefly, this statistic is a measure of the presence of extreme outliers. Genes are median-centered and scaled by the median absolute dispersion. To avoid filtering genes with outliers, if a gene has zero median absolute dispersion, it is instead scaled by the minimum observed non-zero value across the transcriptome. Statistical outliers are identified and their standardized values are summed. Genes classified as variable ranked among the top 400 by this statistic in a given cell type and additionally had an outlier-sum of at least 100. Consistently expressed genes ranked among the top 400 in a given cell type by the Shapiro-Wilk statistic, which measures similarity to a normal distribution, and additionally had a probability of normality greater than 0.005. Functionally enriched categories were identified against matched background sets of expressed genes.

To test the hypothesis that genes with patterns of extreme variation have short RNA half-lives, we used publically available RNA half-life measurements (Schwanhäusser et al., 2011). We classified gene stability following the method used by the original authors, ranking genes by half-

life and categorizing the upper third as slow decaying and the lower third as fast decaying. We tested for an association between categorization of genes as highly variable or consistently expressed and as rapidly or slowly decaying using the chi-square test, rejecting the null hypothesis of no association at $p < 0.05$.

CHAPTER 3: Assessment of single-cell RNA sequencing methods

3.1. Overview

We conducted a large-scale control experiment to assess single-cell RNA sequencing measurements generated by three methods, performing replicate transcriptome amplification and sequencing of reference RNAs at single-cell abundances. Using these data, we provide an assessment of measurement sensitivity, precision and accuracy. All methods detected greater than 70% of the expected number of genes and methods had a 50% probability of detecting genes with abundance greater than 2 to 4 initial molecules. Despite the small number of initial molecules, sequencing depth significantly affected gene detection. While biases in detection and quantification were qualitatively similar across methods, the degree of bias differed, consistent with differences in molecular protocol. Measurement reliability increased with expression level for all methods and we conservatively estimate measurements to be quantitative at ~5-10 molecules. We present several protocol optimizations and list method-specific outlier genes likely to produce problematic measurements.

3.2. Introduction

Single-cell RNA sequencing allows unprecedented resolution for studies of gene expression. Since its introduction in 2009 (Tang et al., 2009), this approach has been used to identify and classify cell types, characterize rare cells, and study gene expression variation across cell populations (Brennecke et al., 2013; Chiu et al., 2014; Deng et al., 2014; Eckersley-Maslin et al., 2014; Park et al., 2014; Piras et al., 2014; Poulin et al., 2014; Sul et al., 2009). In this method, the RNA in a single cell is captured, reverse transcribed to generate cDNA, amplified and sequenced, providing measurements of single-cell transcriptomes with nucleotide-level resolution. Compared with methods to sequence bulk RNA, single-cell RNA sequencing requires additional handling and enzymatic reactions, coupled with substantial molecular amplification.

This combination has the potential to introduce additional experimental errors and molecular biases, which results in the possibility that analytic methods designed for bulk RNA sequencing may not be appropriate for single-cell measurements. Despite substantial methods development over the past six years (Hashimshony et al., 2012; Picelli et al., 2013; Sasagawa et al., 2013; Islam et al., 2014; Jaitin et al., 2014; Lee et al., 2014b), these measurements remain complex and poorly characterized. Though measurement characteristics likely depend substantively on specifics of the experimental protocol used to capture and amplify RNA, there has been limited examination of whether measurement characteristics differ across methods. Additionally, although several common applications of single-cell RNA sequencing rely on measurement sensitivity, there are few assessments of gene recovery and the factors that may affect it.

Here, we characterize expression measurements generated by three single-cell RNA sequencing methods in terms of sensitivity, precision and accuracy. We find that all methods perform comparably overall, but that individual methods demonstrate unique strengths and biases.

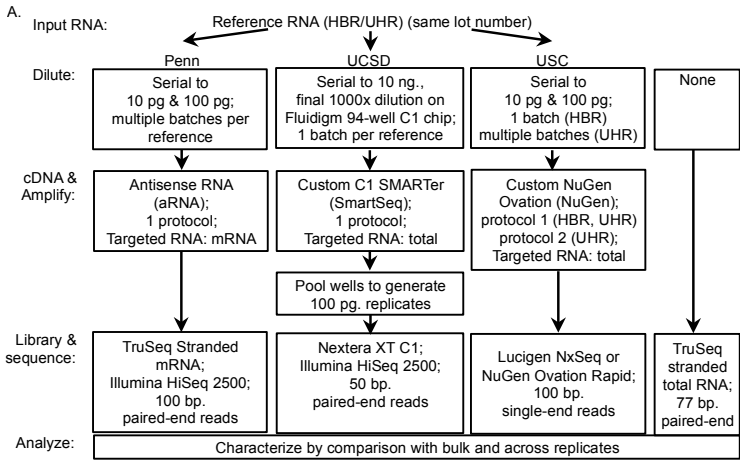
3.3. Results

3.3.1. RNA-sequencing datasets

We performed replicate transcriptome amplifications of Universal Human Reference RNA (UHR) and Human Brain Reference RNA (HBR) that were diluted to near single-cell abundances (10 and 100 picograms (pg.) total RNA or ~200,000 and 2 million mRNA molecules) and were amplified using three single-cell RNA sequencing methods (Figure 3.1a-b). Amplification methods included the antisense RNA protocol (aRNA), a custom C1 SMARTer protocol (SmartSeq) performed on a Fluidigm C1 94-well chip, and a modified NuGen Ovation RNA sequencing protocol (NuGen, Figure 3.1b-c, Table B.1). Bulk UHR and HBR RNA were sequenced directly and served as a reference in analysis. The general experimental scheme was consistent for all dilution replicates; however, there were differences across experimental groups in the specifics of

experimental protocols, necessitated by particular methodologies (Figure 3.1a, see section 3.5.1 and Table B.1 for full details). Because of these experimental differences, head-to-head comparison of methods is not appropriate and our goal is to provide quantitative analyses of factors affecting individual methods, given the dilution and experimental scheme used for the relevant samples. Current results should be used in experimental planning, data analysis, and method optimization rather than as a performance test of any particular method.

Figure 3.1 Experimental design and sequenced data



C.

Single cell protocol	Input (pg)	HBR		UHR		Sequencing							
						Total reads (millions)		% discarded (primer contamin.)		% aligned		% uniquely aligned	
		n	key	n	key	Ave.	Sd.	Ave.	Sd.	Ave.	Sd.	Ave.	Sd.
aRNA	10	18		12		24.3	8.4	0.1	0.1	89.3	6.7	83.1	6.7
	100	3		7		32.0	7.2	0.0	0.0	98.2	1.2	91.1	1.8
C1 SMARTer	10	15		15		14.6	2.0	0.2	0.1	95.0	1.3	78.3	3.4
	100	5		5		14.9	1.8	0.2	0.1	95.4	0.8	78.7	2.4
Nugen Ovation RNASeq V2	10	4		15		24.6	10.9	7.0	11.5	76.3	14.8	63.4	15.9
	100	4		11		26.8	11.2	1.7	2.5	84.6	11.2	73.7	8.9
None	Bulk	3		3		144.3	18.6	0.0	0.0	99.1	0.1	89.3	1.5

A. Dilution experiment summary. See section 3.5.1 for detailed information. **B.** Single cell amplification methods used. Protocols involve two key steps: conversion of RNA (*blue*) to cDNA (*green*), and amplification of cDNA. aRNA targeted poly-adenylated mRNA by using an oligo-dT T7 primer for initial cDNA synthesis. After generating double-stranded cDNA, molecules were amplified using in vitro transcription with T7 polymerase. This amplification procedure was designed to minimize exponential expansion of errors. cDNA generation and amplification were repeated two additional times before library preparation. SmartSeq targeted total RNA using a mixture of poly-T and other primers for initial cDNA synthesis. Full-length transcripts were captured through the template-switching capacity of reverse transcriptase. Double stranded cDNA molecules were amplified using 18 rounds of PCR. All cDNA and amplification reactions were performed on a 94-well Fluidigm C1 chip, intended to reduce experimental variation by performing reactions in small volume. NuGen targeted total RNA through use of proprietary primers for initial cDNA synthesis. Second strand cDNA synthesis was generated using an RNA primer, which was subsequently degraded from the second strand cDNA copy, resulting in linear amplification by DNA replication. This method was designed to minimize exponential amplification of error. **C.** Sample sizes and RNA sequencing statistics by experimental group. Includes color key used in all figures. For analysis based on combined HBR and UHR dilution replicates, solid colors were used. Abbreviations: Human Brain Reference (*HBR*), Universal Human Reference (*UHR*), University of Pennsylvania (*Penn*), University of California San Diego (*UCSD*), University of Southern California (*USC*), picogram (*pg*.), base pair (*bp*.), contamination (*contamin*.), average (*Ave*.), standard deviation (*Sd*.), amplification (*amp*.).

3.3.2. Data processing

Briefly, all samples were aligned to hg19 using STAR aligner (Dobin et al., 2012). Uniquely aligned reads were assigned to Gencode18 gene annotations using HTSeq and htseq-counts (Anders, 2010) and then were depth normalized (Anders and Huber, 2010). Ribosomal genes and genes with short isoforms (<300 nucleotides) were excluded because of differences in sequencing protocols across groups (Figure 3.1a), leaving 42,855 Gencode18 annotated genes for analysis. To avoid artifacts caused by alignment or quantification ambiguities, we also generated a stringently filtered gene list containing 10,039 genes to which reads can be uniquely assigned and referred to these genes as “computationally unambiguous” throughout. Reference

RNA abundances were estimated using bulk UHR and HBR sequencing measurements, aligned and quantified with RSEM (RNA-seq by Expectation-Maximization) (Li and Dewey, 2011). Estimated abundances were concordant with publicly available PrimePCR measurements and with poly-A RNA sequencing measurements (Figure B.1) (SEQC/MAQC-III Consortium, 2014, GEO accession numbers: GPL18522, GSM1362002-GSM1362029, GSM1361974-GSM1362001). The mass of targeted input RNA in diluted replicates was estimated as in (Brennecke et al., 2013) and was used to calculate, for each gene, the expected number of input molecules in a diluted replicate. aRNA selectively targeted poly-adenylated (poly-A) mRNA (Figure 3.1a). We calculated the expected number of input poly-A molecules using publicly available bulk HBR sequencing measurements and in a few instances additionally compare aRNA to these poly-A expectations (SEQC/MAQC-III Consortium, 2014, GEO accession numbers: GSM1362002-GSM1362029). See section 3.5.2 – 3.5.4 for further details.

On average, replicates were sequenced at a depth of 22.0 ± 9.6 million reads (± standard deviation or Sd.). 1.5 ± 5.3 % of reads were discarded due to primer contamination. 89.3 ± 10.6 % of retained reads aligned to the genome, 77.6 ± 11.2 % uniquely (Figure 3.1c). To examine the coverage distribution of each method, we quantified the frequency of mapped reads over several genomic regions of interest (Table 3.1). This distribution differed for the three single-cell amplification methods. The majority of aligned reads for aRNA dilution replicates originated from nuclear exons (excluding rRNA), a substantially larger proportion than that recovered by SmartSeq or NuGen. Nuclear rRNA, including genes, pseudogenes and repeats, comprised a small fraction of reads in all amplified libraries (average ± SD: 0.67 ± 0.65 %); however, mitochondrial rRNA (2 genes) and mitochondrial mRNA (13 genes) did comprise a substantial percentage of reads (average ± Sd.: 16.5 ± 8.4 %). Mitochondrial recovery differed substantially across methods. This difference may translate into a method-specific effect on depth normalization and for this reason mitochondrial genes have been excluded from the subsequent

analyses. The distribution of reads across genomic features also differed substantially across replicates generated by the same method for aRNA and NuGen (Table 3.1).

Table 3.1 Coverage selectivity by method

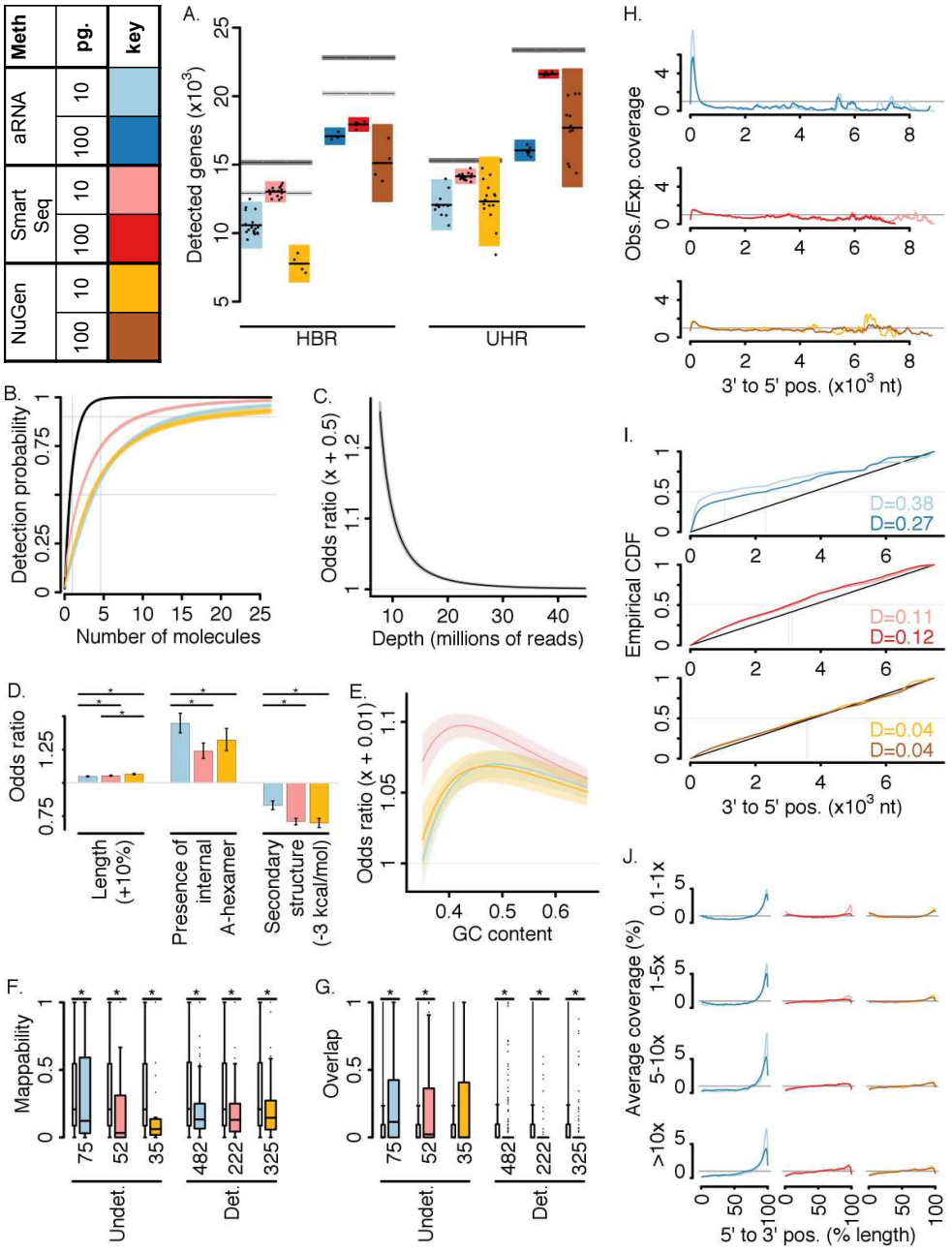
		Genome coverage									
		% exons (excluding rRNA & mitochondria)		% intronic		% rRNA (nuclear)		% rRNA (mitochondrial)		% mitochondrial (non-rRNA)	
Source	Protocol	Ave.	Sd.	Ave.	Sd.	Ave.	Sd.	Ave.	Sd.	Ave.	Sd.
HBR	aRNA	59.07	5.70	23.29	4.68	0.03	0.02	4.82	1.62	12.29	2.79
	SmartSeq	41.00	0.86	39.56	0.85	1.28	0.07	10.02	0.43	7.77	0.26
	NuGen	29.27	4.97	45.09	5.88	1.33	0.40	20.38	5.19	3.65	0.59
	Bulk (Poly-A)	80.13	-	8.52	-	0.08	-	1.96	-	9.01	-
	Bulk (rRNA-depleted)	61.44	0.54	37.89	0.45	0.03	0.01	0.10	0.04	0.25	0.04
UHR	aRNA	71.57	2.34	23.02	2.60	0.06	0.05	1.61	0.19	3.38	0.59
	SmartSeq	35.89	0.89	52.84	0.94	0.31	0.02	6.20	0.26	3.85	0.12
	NuGen	33.00	4.02	39.33	8.99	1.25	0.57	23.47	7.31	2.59	0.62
	Bulk (Poly-A)	86.99	-	7.11	-	0.11	-	0.47	-	5.09	-
	Bulk (rRNA-depleted)	58.17	0.34	41.28	0.29	0.02	0.01	0.03	0.01	0.18	0.02

Average percent of aligned reads assigned to genomic regions for each method. Nuclear rRNA includes rRNA genes, pseudogenes and repeats. See section 3.5.5 for definitions of genomic regions.

3.3.3. Gene detection sensitivity

We calculated the number of recovered genes as a measure of detection sensitivity (Figure 3.2a). All methods demonstrate comparable high gene recovery, recovering greater than 70% of the expected number of genes, with SmartSeq demonstrating the highest recovery (Byar's 95% C.I., Obs. / Exp.: aRNA (0.722, 0.726); SmartSeq (0.877, 0.882); NuGen (0.735, 0.740)). With respect to poly-A RNA, aRNA recovered (0.840, 0.844) of expectation. Variation across samples within each method was substantially larger than expected due to dilution (Figure 3.2a).

Figure 3.2 Single-cell RNA sequencing sensitivity



A. Number of detected genes. Each point represents a single sample. Horizontal black lines indicate group mean. Boxes indicate ± 2 Sd.. Gray horizontal lines indicate 95% CI for the expected number of genes in a diluted replicate, assuming total (*dark gray*) or poly-A (*light gray*) RNA. See section 3.5.6. **B.** Probability of gene detection as a function of the expected number of

input molecules estimated using logistic model (see main text and 3.5.8). Horizontal lines indicate 50% and 90% probability. Vertical lines indicate 1 and at 4.605 molecules (99% probability of ≥ 1 molecule present in diluted replicate). Bands indicate 95% CI. Black line indicates probability of ≥ 1 molecule present in a diluted replicate. **C.** Odds ratio for gene detection as a function of sequencing depth. Horizontal line indicates an odds ratio of one (no gain in detection sensitivity). Band indicates 95% CI. **D.** Odds ratios for differences in biophysical trait values. Error bars indicate a 95% CI. "*" indicates significant difference across pairs of methods (Bonferonni corrected $p<0.05$). **E.** Odds ratio for an increase of 0.01 in GC content. Bands indicate 95% CI. **F.** Boxplot of gene mappability, or the fraction of the gene body that can be aligned to uniquely (see 3.5.7) for computationally ambiguous gene detection outliers (*wide* boxes) and background genes (*narrow* boxes). Both undetected (*Undet*.) and detected (*Det*.) outliers are shown. "*" indicates significant difference (Wilcoxon rank-sum two-way test $p<0.05$). **G.** As in F, but for the fraction of the gene body that overlaps in genomic position with a separate gene annotation. **H.** Nucleotide coverage. Observed over expected coverage normalized for expression level as a function of absolute 3' to 5' position. See section 3.5.10. **I.** Comparison of nucleotide coverage with uniform distribution. Empirical CDF is of normalized per nucleotide coverage. Black diagonal line indicates uniformity. Kolmogorov-Smirnov (KS) statistic for difference from the uniform distribution is in the bottom right, with larger values indicating greater difference between the distributions. **J**. Coverage for genes with different expression levels. Relative 5' to 3' coverage, calculated over 100 equally spaced bins for four expression level categories (rows). See 3.5.10. *Abbreviations:* confidence interval (*CI*); Cumulative distribution function (*CDF*).

Detection of a given gene may depend on parameters such as the input number of molecules, GC-content, presence of internal adenosine monophosphate (A) hexamers, length, strength of molecular secondary structure, and sequencing depth. To assess sensitivity while controlling for these factors, and to estimate the contribution of these factors to gene detection, we fit a logistic regression model to the 10 pg. gene detection data with gene detection as the dependent variable, considering only computationally unambiguous genes to focus on experimental sensitivity. (See section 3.5.7-3.5.8 and Table B.3-Table B.4.) All methods had a 50% probability of gene detection at ~2-4 expected input molecules, controlling for the remaining covariates (Figure 3.2b, Table B.5). We calculated a molecular recovery rate as the predicted probability that a gene with 1 expected input molecule will be detected, scaled by the probability

that at least one molecule of such a gene will be in a diluted replicate. Molecular recovery rates were greater than 0.25 for all methods (95% prediction interval: aRNA (0.262, 0.279), SmartSeq (0.534, 0.558), NuGen (0.315, 0.339)). With respect to poly-A RNA, aRNA recovery rate was (0.320, 0.349).

Despite the small number of total (targeted) RNA molecules a single 10 pg. dilution replicate (estimated here to be ~300,000 molecules), sequencing depth had a highly significant effect on gene detection (Table B.3). Figure 3.2c shows the odds ratio of increasing sequencing depth by 500,000 reads. Briefly, odds indicate the probability that an event occurs (here, that a gene is detected) over the probability that it does not occur. The odds ratio is the relative odds of event occurrence at two values of a covariate, controlling for all other covariates. In this case, for example, an odds ratio of 2 would indicate that an increase of 500,000 sequenced reads doubles the odds of gene detection. The odds of gene detection increased substantially with sequencing depth until a depth of ~15-20 million reads or ~50 reads per input molecule. Here, increasing sequencing depth from 10 to 15 million reads translated into an expected gain of 25.02 % in detected genes. The influence of remaining covariates on gene detection differed across methods (Figure 3.2d-e). The odds of gene detection increased with gene length, and NuGen demonstrated a significantly stronger length effect than aRNA or SmartSeq (Figure 3.2d). The presence of an internal A-hexamer positively influenced the probability of gene detection for all methods, with strongest effect for aRNA. Increased strength of secondary structure decreased the odds of detection for all methods, with significantly smaller effect for aRNA than for SmartSeq or NuGen. While GC content influenced detection probability in a complex manner, SmartSeq demonstrated the strongest GC effect (Figure 3.2e).

A small fraction of computationally unambiguous genes were fit poorly by the logistic model (percent ± 1.96 x binomial standard error: 0.30 ± 0.14 %; see Table B.6 for a list of outliers and section 3.5.9 for details). Each outlier was categorized as “detected” if the gene was

unexpectedly observed and “undetected” if it was unexpectedly missing. Nearly all identified outliers (16/17) were method-specific. A larger proportion of computationally ambiguous genes were poorly fit by the model (3.21 ± 0.23 %, Table B.6), with a sizable minority of outliers (19.81 ± 2.90 %) fit poorly for all methods. These outlier genes had significantly lower fraction of the gene body that could be aligned uniquely than background genes (Figure 3.2f; Wilcoxon rank sum two-way test, $p<0.05$). This was the case for both detected and undetected outliers, indicating that alignment ambiguities increased gene detection uncertainty and generated both false positives and false negatives. Outliers also significantly differed from background in the fraction of the gene body that overlapped in genomic position with a separate gene annotation, with lower overlap among detected outliers and greater overlap among undetected outliers (Figure 3.2g).

To characterize read coverage at the scale of individual base positions, we calculated the observed / expected nucleotide coverage as a function of 3' to 5' position within a gene (Figure 3.2h), normalized such that a uniform distribution of reads along a gene would be assigned a value of one at all positions (see section 3.5.10). Coverage for all methods was significantly different from uniform (Figure 3.2i; Kolmogorov-Smirnov test of equality with uniform distribution $p<10^{-10}$ for all groups); however, NuGen demonstrated the greatest uniformity (Figure 3.2h-i) with similar positional coverage distribution for 10 pg. and 100 pg. dilution replicates. aRNA preferentially covered the 3’ terminal and demonstrated greater 3’ bias for 10 pg. dilution data. SmartSeq showed an intermediate pattern. Gene coverage patterns differed across genes as a function of gene expression level in a similar manner for all methods, with preferential recovery of gene ends for low abundance genes and preferential 3’ coverage for high abundance genes (Figure 3.2j).

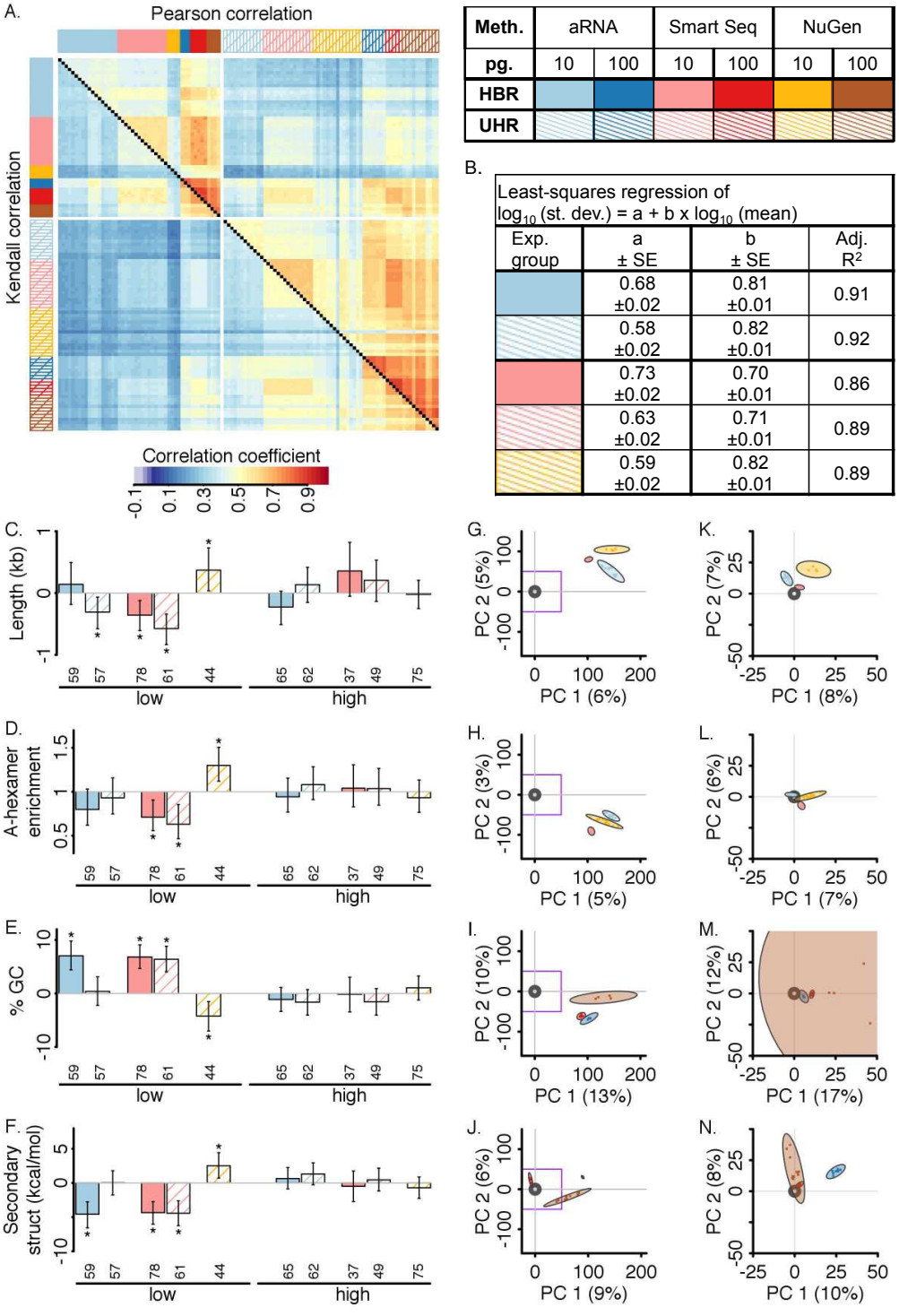
3.3.4. Precision

We next consider the similarity of measurements across dilution replicates, within methods and across methods. Though we cannot quantitatively compare measurement precision

across methods (see Figure 3.1a), the results will be applicable to experimental design and analysis for each method. The average within experimental group pairwise correlation coefficient (± Sd.) was 0.37 ± 0.07 (Kendall) and 0.51 ± 0.09 (Pearson, $\log_{10}$ counts) for 10 pg. replicates and 0.64 ± 0.06 (Kendall) and 0.79 ± 0.06 (Pearson, $\log_{10}$ counts) for 100 pg. replicates (Figure 3.3a; zeros treated as missing values).

To describe the dependence of precision on expression level, we performed least-squares regression of the empirical standard deviation on the empirical mean (both variables log-transformed to satisfy the assumption of residual normality) for 10 pg. experimental groups with sample size >5. The mean was an excellent predictor of standard deviation (Figure 3.3b, adjusted R^2>0.85 and slope coefficient t-test $p<10^{-16}$ in all cases). Genes that were more variable than expected (standardized residuals outside predicted 90% CI) differed little from background in their biophysical characteristics (Figure 3.3c-f), suggesting limited systematic bias in experimental variability. Biophysical characteristics enriched among unexpectedly precise genes with respect to background differed in a method-specific manner (Figure 3.3c-f). For aRNA and SmartSeq, enriched biophysical characteristics were concordant with reduced probability of gene detection (compare Figure 3.3d-e), while NuGen demonstrated the opposite trend. A subset of genes whose standard deviation was poorly predicted by the mean (percent of genes with standardized residuals outside predicted 99.3% C.I.) are listed in Table B.7. We recommend these genes' expression values should be interpreted with caution.

Figure 3.3 Single-cell RNA-sequencing precision



A. Pairwise correlations for all samples. Upper triangle: Pearson correlation. Lower triangle: Kendall correlation. Zeros treated as missing values. Each row and column is an individual sample. Experimental group is indicated by color bars at edge of plot. **B.** Relationship between standard deviation (*st. dev.*) and mean characterized by least squares regression (see section 3.5.11). All estimated coefficients were highly significant (coefficient t-test $p<10^{-16}$). **C-F**. Enrichment of biophysical traits in experimentally precise (*low*) and variable (*high*) genes with respect to background genes (see section 3.5.11). Error bars indicate 95% CI. "*" indicates significant difference ($p<0.05$). Numbers at bottom indicate sample size (number of genes). (C) Median difference in gene length estimated by Hodges-Lehman statistic. Significance: Wilcoxon rank sum two-way test. (D) Relative risk of containing an internal A-hexamer. Significance: Fisher's exact test. (E) As C for % GC content. (F) As C for strength of local secondary structure. **G-J.** PCA projection of dilution data on PC 1 and 2. Plots were centered so that bulk UHR or HBR was positioned at the origin. Points represent individual dilution replicates. Colored ovals represent bivariate normal 95% confidence ellipses. % Sd. explained by a PC is indicated in axis label. See 3.5.11. (G) HBR 10 pg. (H) UHR 10 pg. (I) HBR 100 pg. (J) UHR 100 pg. **K-N.** As G-J, but using only abundantly expressed genes (see main text). Axis scales differ from G-J, with axes in equivalent to the purple-boxed region in G-J. (K) HBR 10 pg. (L) UHR 10 pg. (M) HBR 100 pg. (N) UHR 100 pg. *Abbreviations*: Standard error (*SE*); confidence interval (*CI*); kilobases (*kb*); kilocalories (*kcal*); mole (*mol*); principal components analysis (*PCA*); principal component (*PC*); standard deviation (*sd*).

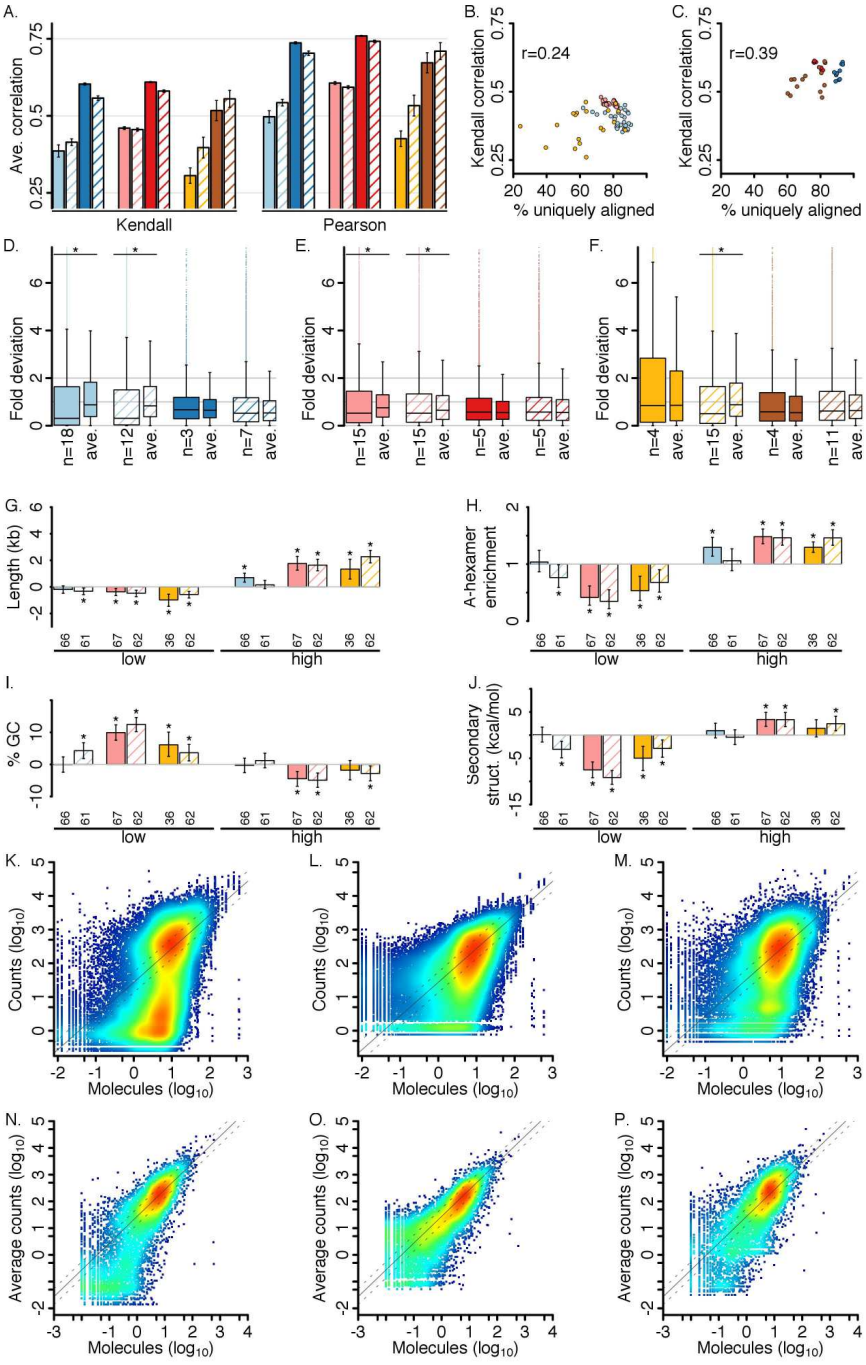
Separate principal components analysis (PCA) of each HBR and UHR for 10 pg. dilution data demonstrated that variation between single cell and bulk measurements predominate over differences between single cell methods (Figure 3.3g-h); however, dilution replicates segregated by amplification methods along the first two principal components (PCs) indicating differences across methods in the multivariate covariance structure of experimental variation. Differences across methods were also apparent for 100 pg. dilution replicates (Figure 3.3i-j), and, though these measurements were more similar to bulk measurements, differences between dilution replicates and bulk measurements persisted. Single cell methods demonstrated differences in gene detection sensitivity and bias (Figure 3.2), which may contribute substantially to multivariable structure of experimental variation. We repeated PCA on a subset of genes with

greater than 18.5 expected input molecules (expected probability of detection for “typical” gene > 0.9 for all methods). On highly abundant genes, dilution replicates were substantially more similar to bulk measurements (Figure 3.3k-n) and differences across methods were substantially smaller. However, in all cases, the within method pattern of covariation (direction of ellipses) and the bias dispersal around the bulk expected value (position of the centroid of the ellipses) differed for both source RNA and individual methods.

3.3.5. Accuracy

We calculated pairwise correlation coefficients of diluted replicate measurements with bulk HBR or UHR measurements as a metric of overall accuracy (Figure 3.4a). For this and the below evaluations of the accuracy of gene measurements in individual dilution replicates, only non-zero gene counts were considered in order to focus on quantitation rather than sensitivity. 10 pg. dilution replicates demonstrated an average pairwise correlation with reference (± 1.96 x s.e.m.) of 0.42 ± 0.01 (Kendall) and 0.55 ± 0.01 (Pearson, $\log_{10}$ counts). 100 pg. replicates showed greater similarity with reference (0.57 ± 0.01 (Kendall) and 0.72 ± 0.01 (Pearson, $\log_{10}$ counts). Correlation with reference had a modest association with percent unique alignment (Figure 3.4b-c).

Figure 3.4: Single-cell RNA sequencing accuracy



Meth.	aRNA		Smart Seq		NuGen	
pg.	10	100	10	100	10	100
HBR						
UHR						

A. Average pairwise correlation of diluted replicates with bulk HBR or UHR. Zeros treated as missing values. Error bars indicate ± 2 s.e.m.. **B-C.** Relationship between % unique alignment and similarity with bulk HBR or UHR. "r" indicates Pearson correlation of x and y. (B) 10 pg. (C) 100 pg. **D-F.** Distribution of fold deviation across genes. *Wide* boxes represent measurements in individual replicates. *Narrow* boxes represent average measurements across replicates. Y-axis was truncated for visualization and 99%ile values for wide boxes are in panel descriptions below, ordered to match plot. "*" indicates significant difference between individual and average measurements (Wilcoxon rank sum test of greater fold deviation in average measurements, $p<0.05$). (D) aRNA. 99%ile values: 2066; 1442; 514; 718. (E) SmartSeq. 99%ile values: 877; 770; 553; 598. (F) NuGen. 99%ile values: 2332; 1766; 784; 937. **G-J.** Enrichment of biophysical traits among underestimated (*low*) and overestimated (*high*) genes with respect to remaining genes. See section 3.5.12. Plot notation and statistics are as in Figure 3.3C-F. (G) Median difference in gene length. (H) Relative risk of containing an internal A-hexamer. (I) As G, for % GC content. (F) As G, for strength of local secondary structure. **K-M.** Density scatter plots of normalized read counts in individual 10 pg. replicates vs. expected number of input molecules. See section 3.5.12 for method. Red indicates high density. Solid line indicates expected read count and hashed lines indicate ± 2 fold. (K) aRNA. (L) SmartSeq. (N) NuGen. **N-P.** Density scatter plots of average normalized read counts vs. number of input molecules. (N) aRNA. (O) SmartSeq. (P) NuGen. *Gene filtering*: D-F and K-P considered computationally unambiguous genes and excluded gene detection outliers. D-J considered genes with greater than 95% probability of presence in a diluted replicate. *Abbreviations*: confidence interval (*CI*); kilobases (*kb*); kilocalorie (*kcal*); mole (*mol*).

To assess the accuracy of individual gene estimates, we calculated the fold deviation of normalized read counts with respect to bulk HBR or UHR measurements (or the observed over expected read count; Figure 3.4d-f, section 3.5.12). For all methods and input amounts, the median fold deviation was less than 1 and a subset of genes was extensively overestimated. The average read count across 10 pg. replicates demonstrated improved median accuracy relative to individual measurements (fold deviation closer to 1), except in the case of NuGen HBR replicates (Figure 3.4d-f). To identify sources of systematic bias, we compared biophysical characteristics of overestimated and underestimated genes (genes with median fold deviation in the top or bottom 5%ile) with those of background genes (Figure 3.4g-j). Overestimated genes were substantially

longer than remaining genes and more frequently contained an internal A-hexamer. For NuGen and SmartSeq, these genes also had lower GC content and weaker local secondary structure than remaining genes. Underestimated genes demonstrated the opposite tendencies: compared to background genes, they were shorter, less frequently contained internal A-hexamers, had higher GC content and stronger secondary structure than background. Overall, aRNA demonstrated less systematic bias than NuGen or SmartSeq. Highly inaccurate genes (genes with median fold deviation in the top or bottom 1%ile) are catalogued in Table B.8.

Smoothed density scatter plots demonstrated method-specific transfer functions between the expected number of input molecules and the number of read counts in an individual replicate (Figure 3.4k-m). This relationship was roughly linear at expression levels greater than ~5-10 expected input molecules. At low to mid expression levels measurements were frequently underestimated, and this produced an expanded range of measured abundances with respect to input abundances at these expression levels, particularly for aRNA and NuGen. This effect was reduced for average expression levels (Figure 3.4n-p).

3.3.6. Protocol variations

We evaluated the effects of several protocol variations on measurement quality (Table 3.2). The aRNA protocol used for the primary analysis includes cDNA purification before initial amplification, and 3 rounds of IVT amplification followed by dilution of amplified cDNA before library preparation (Figure 3.1b). Elimination of initial cDNA cleaning significantly improved sensitivity and accuracy, as did reduction to two rounds of IVT amplification and elimination of dilution prior to library generation (Table 3.2). An optimized protocol incorporating both changes, demonstrated substantial improvements in the number of detected genes and pairwise correlation with the bulk (Table 3.2).

The addition of ERCC spike-in transcripts provides an internal control (Jiang et al., 2011); however, the addition of a large amount of synthetic RNA to a sample may decrease biological

sensitivity. We found no significant difference in sensitivity, precision or accuracy across matched dilution replicates with and without the addition of ERCCs (Table 3.2). Individual ERCC transcripts were found to be problematic for SmartSeq and aRNA in a method specific manner (Figure B.2).

Strand-specific RNA sequencing may improve detection sensitivity and reduce false positive detection. Stranded quantification of aRNA replicates detected slightly fewer genes than non-stranded quantification; however, it also detected significantly fewer genes that were not observed in the bulk, and genes that were detected only by stranded quantification were supported by significantly more reads than genes detected only by non-stranded quantification (Table 3.2).

Table 3.2 Evaluation of protocol variations

Comparison of dilution replicates generated using modified protocols with control dilution replicates. Sample information can be found in Table B.2 and protocol information in section 3.5.13. # genes detected only considers genes observed in bulk HBR or UHR. Kendall correlation was calculated excluding zeros in either sample. Unpaired comparisons were made using Wilcoxon two-way rank sum test for difference in medians. Paired comparisons were made using Wilcoxon two-way rank sign test for difference in medians. The null hypothesis of no difference was rejected at $p<0.05$. Median difference between groups, with 95% CI, was calculated using the Hodges-Lehman statistic.

			Modified group		Control group		Median difference (Modified:Control)			Test		
Variation	Category	Trait	Sample size	Median	Sample size	Median	Statistic	95% C.I.		Paired/ Unpaired	Statistic	p-value
No initial cDNA cleaning	Sensitivity	# genes detected	6	11072	6	10249	782.5	293.0	1678.0	Unpaired	36	0.002
	Precision	Pairwise correlation across samples	15	0.344	15	0.306	0.039	0.028	0.051	Unpaired	215	1.79E-06
	Accuracy	Pairwise correlation with reference	6	0.400	6	0.383	0.019	0.008	0.049	Unpaired	36	0.002
Reduce rounds of cDNA amp.	Sensitivity	# genes detected	5	11936	3	11062	1051.0	33.0	4122.0	Unpaired	15	0.036
	Precision	Pairwise correlation across samples	10	0.329	3	0.333	-0.006	-0.034	0.018	Unpaired	12	0.692
	Accuracy	Pairwise correlation with reference	5	0.430	3	0.401	0.029	0.009	0.066	Unpaired	15	0.036
Optimized aRNA	Sensitivity	# genes detected	5	11936	18	10286	1810.0	942.0	3354.0	Unpaired	84	0.002
	Precision	Pairwise correlation across samples	10	0.329	153	0.306	0.019	0.002	0.035	Unpaired	1084	0.028
	Accuracy	Pairwise correlation with reference	5	0.430	18	0.377	0.055	0.028	0.083	Unpaired	79	0.009
Add ERCCs	Sensitivity	# genes detected	5	10154	4	10101	138.5	-397.0	798.0	Unpaired	13	0.556
	Precision	Pairwise correlation across samples	10	0.287	6	0.282	0.002	-0.012	0.013	Unpaired	34	0.713
	Accuracy	Pairwise correlation with reference	5	0.364	4	0.369	-0.003	-0.016	0.019	Unpaired	9	0.905
Perform strand-specific sequencing	Sensitivity	# genes detected	17	10006	17	10325	-267.0	-331.5	-218.5	Paired	0	1.53E-05
	Sensitivity	Depth of unique genes	17	6.397	17	0.848	5.46	3.68	6.75	Paired	153	1.53E-05
	Sensitivity	# genes not in bulk	17	534	17	610	-62.0	-76.0	-49.5	Paired	0	1.53E-05
	Accuracy	Pairwise correlation with reference	17	0.362	17	0.376	-0.013	-0.017	-0.009	Paired	2	4.58E-05

3.4. Discussion

In light of these results, we briefly discuss a few topics related to experimental planning, method optimization and data analysis.

Though the data presented here are not appropriate for use in head-to-head methods comparisons, some results may be helpful in selecting an appropriate method. The enriched coverage of exons in aRNA may be beneficial for studies of mRNA, and the retention of transcript strand information is unique to aRNA at this point. SmartSeq and C1 microfluidic device generates reproducible replicates and high detection sensitivity, presumably due to more uniform liquid handling and retention of material due to lack of vessel transfer. The uniformity of coverage provided by NuGen (Figure 3.2h-i) may be beneficial studies of isoform use and splicing. We note that, in our hands, NuGen reactions were inconsistent and we had repeated amplification failures with this method.

In selecting sequencing depth, there is a trade-off between gene detection sensitivity and cost. Typically, a small number of genes comprise the bulk of RNA molecules in a transcriptome. Sequencing at low depths should be sufficient to reproducibly detect and quantify these abundant genes; however, the majority of genes in a typical transcriptome are low abundance. Because of this, the number of genes detected depends heavily on sequencing depth (Figure 3.2c). Our results suggest that sequencing deeply may produce richer transcriptome measurements, and that the extent of multiplexing used in single-cell RNA sequencing experiments should be considered carefully in the context of a specific experimental plan.

Alignment and expression quantification require assumptions about the process by which sequenced reads are sampled from the transcriptome. Many bioinformatics tools rely on assumptions that, while appropriate for bulk RNA sequencing, may not be appropriate for single cell data. Our results suggest that this may be the case: PC analysis demonstrated greater

difference between single cell and bulk measurements than across single cell methods (Figure 3.3g-j). Additionally, this sampling process likely differs across single cell methods. Results above demonstrate differences across methods in characteristic per nucleotide coverage patterns (Figure 3.3h-i), in experimental covariance across genes (Figure 3.3g-j), in the transfer function between the number of input molecules and the measured read counts (Figure 3.4m-o), and in the qualitative relationship between the probability of gene detection and the number of input molecules (compare the shape of curves in Figure 3.2b across methods). Efforts to develop quantification and normalization algorithms for single cell RNA sequencing (as in (Brennecke et al., 2013; Grun et al., 2015; Kharchenko et al., 2014)) in a method-aware manner will increase the biological impact of this method.

Missing values due to lack of sensitivity and the presence of large valued outliers may cause complications for depth normalization methods. Large variation across samples and substantial differences across methods in the fraction of reads assigned to mitochondrial RNA (Table 3.1) will propagate to sample and method differences in relative read counts. More generally, we observed large variation in the distribution of reads across broad genomic regions (Table 3.1). Because each genomic region accounted for a substantial number of reads and input molecules, the observed differences across methods and within methods cannot be simply explained by sampling error. Similarly, variation across samples in the number of detected genes cannot be easily explained by dilution (Figure 3.2a). This behavior might be explained by global differences in reaction efficiencies across samples, as suggested previously (Grün et al., 2014b); however, the experimental sources of such differences in a controlled experiment are unclear.

Some bioinformatics challenges might be reduced through experimental optimizations, for example by increasing detection sensitivity and reducing amplification biases. Limiting cleaning, reducing the extent of amplification required, and limiting sample dilution may be

productive avenues. Methods to experimentally deplete highly abundant and variably recovered mitochondrial RNA, if not of experimental interest, may also be of use.

We found subsets of genes to be problematic for gene detection and accuracy, and to demonstrate unusual variability or precision, in a method-specific manner (Table B.6-Table B.8). We recommend that genes on these lists be treated with caution, filtered before analysis or interpreted with care. We similarly found several ERCC spike-in transcripts to be problematic (Figure B.2), and recommend selecting a subset of reliable ERCC transcripts for use as reference measurements.

All examined methods demonstrated good gene recovery and a linear relationship between input molecular abundances and measured expression levels at mid- to high-expression levels or greater than ~5-10 input molecules. This corresponds to ~4,000–8,000 reliably measured genes for the reference transcriptomes examined here. This level of resolution for gene expression measurements has and will continue to facilitate biological discovery.

3.5. Methods

3.5.1. Experimental design

Each collaborating center obtained reference RNA with the same lot number for each of two reference sources (Universal Human Reference (UHR) RNA and Human Brain Reference (HBR)) and performed replicate amplification using a single amplification method, detailed below. These methods relate to Figure 3.1.

SmartSeq: Reference RNA was diluted to an intermediate stock solution by serial dilution. A final 1000-fold dilution occurred on the C1 chip, such that individual wells in a given batch contained 9.99 pg. sampled from a common intermediate dilution. ERCC spike-in RNA was also added for a final mass of approximately 7 femtograms (fg.) per sample, a 4,000,000x dilution from stock. Samples for each source RNA were prepared in single batches. After amplification,

cDNA from the entire C1 96-well plate was quantified using picogreen. C1 chips with an average yield of less than 3 nanograms were discarded. The top 15 reactor wells by cDNA concentration were selected as representative 10 pg. samples for sequencing library preparation. Another 50 wells were selected by the same criteria. These were pooled in sets of 10, generating 5 100 pg. samples for each HBR and UHR. All samples for a given source were prepared in a single sequencing library preparation batch using Nextera XT C1 protocol.

NuGen: HBR samples were prepared in a single batch using amplification protocol 1, generating 4 10 pg. and 4 100 pg. amplified replicates. UHR samples were prepared in two batches, using either amplification protocol 1 or 2, generating 15 10 pg. and 11 100 pg. samples. Sequencing library preparation was performed for each batch of samples using either Lucigen NxSeq or NuGen Ovation Rapid protocol (see Table B.1).

aRNA: Amplification was performed as described in (Morris et al., 2011). HBR samples were prepared in 4 batches from separate dilutions of reference RNA, generating 19 10 pg. and 3 100 pg. amplified replicates. ERCC spike-ins were added to 5 of the 10 pg. replicates before amplification at a dilution of 4,000,000x from stock. UHR samples were diluted and amplified in 2 batches from separate dilutions of reference RNA, generating 12 10 pg. and 7 100 pg. amplified replicates. Sequencing library preparation was performed using Illumina TruSeq Stranded mRNA protocol modified to begin with amplified aRNA. A small numbers of reads were assigned to ERCC transcripts in replicates from the batch where ERCCs had been added that did not have spike-ins added (average of 0.5% of the number of reads assigned in spiked samples). 18 additional HBR 10 pg. replicates were amplified using aRNA for protocol optimization experiments (see Table B.2). These samples were treated separately and were excluded from primary analysis.

Bulk UHR and HBR: For each reference RNA, three sequencing libraries were generated from bulk material at the same laboratory as the SmartSeq replicates. Cytoplasmic and

mitochondrial ribosomal RNA (rRNA) were depleted using Ribo-Zero Gold as part of Illumina TruSeq Stranded Total RNA protocol. Samples were sequenced on Illumina HiSeq 2000. We accessed publicly available bulk sequencing of HBR and UHR generated using poly-A selected RNA generated using standard Illumina mRNA-Seq protocol and sequenced on Illumina HiSeq 2000 using 100 bp. paired-end reads. (SEQC/MAQC-III Consortium, 2014,GEO accession numbers: GSM1362002-GSM1362029 (HBR), GSM1361974-GSM1362001 (UHR), downloaded in May 2015.) These samples were generated as part of a larger experiment to evaluate bulk RNA sequencing where poly-A sequencing was performed at seven sites. For each HBR and UHR, four replicate libraries generated at the NYG site were used. Sequenced read data for each source were pooled. We additionally used publicly available PrimePCR measurements generated by the SEQC/MAQC-III Consortium using UHR and HBR RNA (SEQC/MAQC-III Consortium, 2014, GEO accession number: GPL18522, downloaded in Feb. 2015) to evaluate our reference gene abundance estimates.

Because of differences in experimental design, direct comparison across methods of precision and the effect of input RNA abundance is difficult. For example, input RNA amount as a factor have different meanings for the different amplification methods: for SmartSeq, because 100pg samples were constructed by pooling 10 pg. samples after cDNA amplification, any resulting effects involve library construction, while for aRNA and NuGen resulting effects reflect both cDNA amplification steps and library steps.

3.5.2. Alignment and quantification

These methods relate to section 3.3.2. Low confidence nucleotides (with Phred score less than 20) were treated as unknown and replaced with Ns. Unknown nucleotides (Ns) at the ends of reads were trimmed. Poly-A and method-specific adapter sequences were trimmed from the 3' end of reads using in-house software (Fisher and Kim, 2015). Reads were aligned to the human reference genome, build hg19, and to ERCC spike-in transcript sequences using STAR

(Dobin et al., 2012). We retained reads that aligned to at least 40% (paired-end) or 60% (single-end) of trimmed length or 30bp, whichever was greater. In addition, we discarded reads with greater than 30% mismatched positions in trimmed length. Uniquely aligned reads were assigned to Gencode18 gene annotations and to ERCC transcripts using HTSeq and htseq-counts. Reads overlapping multiple annotations were assigned to a single gene or discarded using the intersection non-empty method (Anders et al., 2015). We normalized raw read counts for differences in sequencing depth using size-factors estimated by the method proposed by Anders and Huber and implemented in DESeq (Anders and Huber, 2010) after filtering genes as described in section 3.5.3, below. aRNA sequenced data retained RNA strand information, but we did not use this information in quantification so that that all samples were consistently handled. For protocol optimization analysis (Table 3.2), aRNA samples were re-quantified using strand information where applicable. Each method demonstrated different dependence of read counts on gene length (Figure 3.2h) and so no single length normalization procedure was appropriate. Analysis was completed without length normalization.

To estimate input RNA abundances, raw sequencing data from all three ribosome-depleted bulk HBR or UHR replicates were pooled resulting in a single sample for each HBR and UHR with sequencing depth of ~400 million reads. Sequencing characteristics of bulk RNA sequencing are relatively well known and we used a model theoretic method to estimate reference gene expression. We used RSEM (RNA-seq by Expectation-Maximization, version 1.2.18, using Bowtie version 1.1.1) strand-specific quantification (Langmead et al., 2009; Li and Dewey, 2011). Poly-A tails were not added to transcripts. RSEM gene abundances were normalized to transcripts per million (*TPM*). 50.4% and 51.1% of reads aligned to genes for HBR and UHR, respectively.

We validated the robustness of the RSEM abundance estimates by comparing them to estimates generated using two additional algorithms. First, we used HTSeq and htseq-counts

(Anders, 2010) in the intersection non-empty mode as described above. This method makes few assumptions about the distribution of sequencing reads along transcripts. Second, we used a modified version of Maxcounts (Finotello et al., 2014), a method designed to be robust to differences in sequencing protocol. In the modified version, each gene was assigned the 95%ile depth of coverage value across covered exons. For both HTSeq and Maxcounts, quantification was strand-specific and estimates were normalized to reads per million (*RPM*). Counts were also compared to PrimePCR measurements (see section 3.5.1). To compute gene abundance estimates using PrimePCR, we removed undetectable genes (CT>35, based on a CT of 35 for one DNA molecule (SEQC/MAQC-III Consortium, 2014)) and then subtracted 35 from each gene's CT value to generate $\log_2$ number of molecules, which were then converted to $\log_{10}$ units. Genes with multiple reported CT measurements were removed, leaving 11,788 (UHR) and 11,572 (HBR) gene measurements for analysis. Pairwise scatter plots and correlations can be found in Figure B.1. All quantification algorithms provide similar estimates. We used RSEM quantification throughout because this method provides isoform expression level estimates, which allow more fine-tuned estimates of gene characteristics (such as GC content and length).

Ribosomal and mitochondrial RNA were depleted from bulk HBR and UHR samples (see section 3.5.1). We compared estimated RNA abundances based on these samples to abundance based on samples generated using poly-A RNA to determine whether the method of RNA selection substantively affected abundance estimates. Expression estimates were similar across library preparation methods and the library generated with ribosomal and mitochondrial depletion demonstrated the greatest similarity with qPCR measurements (Figure B.1b). RSEM expression level estimates based on ribosomal and mitochondrial RNA depleted samples were used as "truth" throughout.

3.5.3. Excluded and unambiguous genes

These methods relate to section 3.3.2. We excluded ribosomal genes, genes with short isoforms, and genes on the mitochondrial chromosome, as described in the main text. Inferences made by bioinformatics methods may affect sensitivity, precision, and quantification accuracy for any individual gene. We identified a stringent set of genes to which reads could be uniquely aligned, in order to focus on sensitivity, precision and accuracy of the molecular measurements. Identified genes did not overlap in genomic positions with exons from any other annotated gene on either strand and could be aligned to uniquely across the entire gene. As a measure of mappability we used the Gencode CRG Alignability track for reference genome hg19, generated by the ENCODE project and downloaded as a bigwig file from the UCSC Genome Browser on Sept. 23 2014 (Kent et al., 2002). This track contains sliding windows of k-mers and a record of how many locations in the genome each k-mer aligns using the GEM aligner allowing up to two mismatches. We used k equal to 50 nucleotides because the minimum read length in this study was 50 base pairs. Genes where all sliding windows align to only one location were considered uniquely alignable.

3.5.4. Expected number of molecules in diluted replicates

These methods relate to section 3.3.2. We estimated the expected number of molecules in a diluted replicate in three steps. First, we estimated the fraction of total input RNA that was targeted for cDNA synthesis and used this to find the mass of targeted RNA. Second, we converted this mass to a total number of input molecules using the average transcript length for each HBR and UHR. Third, we converted gene relative expression levels to expected numbers of molecules in a diluted replicate.

To estimate the mass of RNA targeted for cDNA synthesis, we followed the method proposed in (Brennecke et al., 2013). For each SmartSeq dilution replicate, we calculated the percent of reads assigned to ERCC transcripts, with respect to the total number of reads

assigned to genes that were retained after filtering. (SmartSeq samples were used because all replicates included ERCC spike-ins.) We divided the known ERCC mass (7.12 or 71.2 femtograms) by the average percentage of reads assigned to ERCC transcripts to get the total mass of targeted transcripts and ERCC molecules and therefore the mass of targeted transcripts. By this method, we estimated the following masses for targeted molecules: 0.24 pg. (HBR 10 pg. replicates), 2.4 pg. (HBR 100 pg. replicates), 0.26 pg. (UHR 10 pg. replicates) and 2.6 pg. (UHR 100 pg. replicates). The average transcript length for HBR, based on RSEM relative gene expression level estimates, was 1,535.56 nucleotides (average transcript mass of 8.175×10^{-7} pg. and 288,600 molecules in 10 pg. replicate); for UHR, it was 1,348.39 nucleotides (average mass of 7.179×10^{-7} pg. and 364,762 molecules in a 10 pg. replicate). For ERCC molecules, the expected number was calculated directly from the known mass of spiked-in materials and the known molarity of each spike-in transcript. We repeated this analysis using five aRNA HBR 10 pg. samples that contained ERCC spike-ins to estimate the mass of targeted mRNA in a diluted HBR replicate. The mass of targeted mRNA was estimated to be 0.15 pg. (HBR 10 pg. replicates). We used RSEM relative gene expression level estimates for poly-A selected bulk HBR samples to estimate the number of targeted mRNA molecules. The average mRNA transcript length in HBR was estimated to be 1,968.73 nucleotides (average mass of 1.048×10^{-6} pg. and 143,631 molecules in a 10 pg. replicate).

3.5.5. Genomic distribution of sequenced reads

These methods relate to Table 3.1. Genomic regions were assigned to eight categories hierarchically so that each region was assigned to only one category and so that each read was greedily categorized in the following order: rRNA exon, rRNA repeat, exon (excluding rRNA), intron, flank, intergeneic. Regions were defined based on the following annotations. Exons and introns were assigned based on Gencode18 annotations. Flanks were assigned to 5 kilobases up- and down-stream from gene terminals. rRNA refers to Gencode18 annotations with "rRNA" as the gene_type, which includes 5S pseudogenes. rRNA repeat refers to RepeatMasker

annotations for the rRNA class of repeat. RepeatMasker annotations for reference genome hg19 were downloaded from UCSC table browser as a gtf file from the UCSC genome browser on June 23, 2015. Remaining regions were classified as intergenic. Primary alignments for all reads, including multimapping reads, were assigned to these regions using htseq-counts (Anders et al., 2015). The STAR aligner assigns a single primary alignment to each read, with multi-mapping reads assigned the alignment with the best alignment score, if only one such alignment exists, or a randomly selected alignment from the set of best alignments. (Multi-mapping reads were included for this analysis because many rRNA regions demonstrate substantial similarity such that it was difficult to uniquely align reads to these regions.) Haplotype and random chromosomes were excluded.

3.5.6. Number of detected genes

These methods relate to Figure 3.2a. Genes not observed in the bulk were ignored. The expected number of genes in a diluted replicate was calculated as follows. We assumed that the number of molecules in a tube for a given gene are Poisson distributed with mean equal to the expected number of input molecules and that genes are independent. The presence or absence of a given gene follows a Bernoulli distribution, with the probability of success equal to the probability that at least one molecule for the gene is in the diluted replicate. The number of genes in a diluted replicate is then drawn from a Poisson-Binomial distribution. We used the R package poibin to find a 95% CI for the expected number of genes in a diluted replicate. We performed simulations of the dilution experiment to check robustness of the result to violation of the independence assumption. Simulation results matched theoretical results (data not shown). We performed this analysis both assuming that total RNA was targeted for capture and assuming mRNA was targeted for capture (see section 3.5.4, above). Because UHR aRNA dilution replicates did not contain ERCC spike-ins, we could only estimate mRNA expectation for HBR.

3.5.7. Gene traits

These methods relate to Figure 3.2, Figure 3.3 and Figure 3.4. We compiled a set of gene characteristics for use in bias exploration. Traits calculated include GC content and length, both known sources of bias for bulk RNA sequencing (Zheng et al., 2011). Poly-T priming was used by aRNA and SmartSeq and may introduce a bias for genes with internal stretches of adenosines, and so we also computed the presence or absence of an internal A-hexamer (6 or more sequential As). RNA secondary structure may hinder biochemical reactions and we assigned a score for the average strength of local secondary structure. To do this, we calculated the minimum free energy predicted by Vienna RNAFold (version 1.7.2) (Lorenz et al., 2011) for 100 nucleotide-sliding windows along the length of each isoform (step size of 1 nucleotide) and reported the average across all windows. All traits are calculated based on Gencode18 annotated isoforms. Genes were assigned the average of isoform traits, weighted by the relative expression level of isoforms estimated by RSEM quantification of bulk HBR or UHR. We also calculated two metrics of bioinformatics complexity for each gene. As a measure of alignment complexity, we calculated the fraction of 50 base pair windows that were reported to be uniquely alignable in the Gencode CRG Alignability track (Derrien et al., 2012) (see section 3.5.3). As a measure of quantification complexity, we calculated the fraction of the gene body that overlaps with another annotation on either strand. Both of these metrics were calculated over the union of exons for each gene.

3.5.8. Detection logistic regression

These methods relate to Figure 3.2. For model fitting, we used computationally unambiguous genes (see section 3.5.3) that were observed in bulk HBR or UHR. Genes within the upper or lower 2.5%ile value for any biophysical trait were excluded so that covariate ranges were well sampled. After filtering, 5,645 genes were included in analysis. The analysis was performed on 10 pg. dilution replicates. 100 pg. dilution replicates were not included because of

the small sample size of these groups and because of differences between groups in how these dilution replicates were generated (see Figure 3.1a and section 3.5.1). A single model was fit containing both HBR and UHR dilution replicates, in order to increase sample size and simplify analysis. A random 90% of the data were used in model development and fitting, with the remaining 10% used to assess model fit. Final sample size for model development was 323,194 observations and for validation it was 45,486 observations.

To determine the best parametric form for each independent variable we followed the multivariate fractional polynomial method. In brief, this method (developed by Royston & Altman, 1994) searches a small range of possible polynomial functions of each independent variable to identify the transform that results in the best model, defined as having the largest log-likelihood. Both one- and two-term transforms can be tested. Before selection of a "best" transform, fit models using transformed variables are compared to the linear case (and to each other, if both a one- and two-term transformation are considered) using a likelihood ratio test (here the null hypothesis of no difference in fit was rejected at $p<0.001$). See Hosmer et al. for more details (Hosmer, Jr. et al., 2013). In a multivariate case, transformations are tested on individual covariates iteratively in the context of the multivariate model in order of decreasing significance, retaining selected transformations for previously tested covariates. Once all variables have been tested the process repeats, beginning with the previously identified best transforms, until no additional changes are significant. We used a closed test procedure for determining significance (see Hosmer et al.), permitting two-term transformations for the number of molecules and GC content. Single-term transforms were permitted for gene length, strength of local secondary structure and sequencing depth for the sake of model simplicity and interpretability. We used the R mfp package for this analysis (Ambler (original) and Benner (modified), 2015). For selecting parametric form, all samples were treated together, ignoring amplification method. By this method, the selected model is:

$$\text{Logit}(E(Y|M,L,G,S,A,D)) = \beta_0 + (\beta_1\sqrt{M}) + (\beta_2 \log(M)\sqrt{M}) + (\beta_3 \log(L)) + (\beta_4 G^{-2}) + (\beta_5 \log(G) G^{-2}) + (\beta_6 (S+39.1)/10) + (\beta_7 A) + (\beta_8 (D/10)^{-2})$$

where M represents the expected number of input molecules in a diluted replicate, L represents the gene length (in kilobases), G represents the gene GC content, S represents the gene strength of local secondary structure (kcal/mol, shifted and scaled for stability), A indicates the presence of an A-hexamer within the gene body, and D represents the sequencing depth (per million reads, scaled for stability).

In the final model, amplification method was encoded as dummy variables so that method-specific coefficients were found for all independent covariates, with the exception of sequencing depth. We fit a single coefficient for depth across all methods to increase the covariate range. The final model was fit excluding 17 large influence genes (having Cook's Distance >0.001 for at least two observations in each of at least two methods) using R built-in glm function with family (error model) set to binomial (R Development Core Team, 2010). The final model can be found in Table B.3. Model fit was assessed using normalized Chi-Square (proposed by Osius and Rojek) and normalized Sum-of-Squares goodness-of-fit statistics, evaluated on a random 10% of the data excluded from model development (Table B.4, and see Hosmer et al. for details). To assess the benefit of including biophysical and sample covariates, in addition to the expected number of input molecules, we calculated the area under the receiver operating characteristic curve (AUC) for classification using the model, and separately for classification based on the expected number of input molecules alone. AUC provides a measure of the probability that the classifier will assign a higher score to a randomly selected detected gene than a randomly selected undetected gene. AUC average and standard deviations were calculated over 10,000 bootstrap replicates. To determine whether the model was sensitive to read length or paired end status, we calculated fit statistics for data truncated *in silico* to 50 base pair single-end reads (Table B.4). We additionally tested extension of model to ERCC spike-in

molecules (using SmartSeq and aRNA 10 pg. dilution replicates containing spike-ins) and to dilution samples beginning with 100 pg. input RNA (Table B.4). For these additional validations, a random 5,000 observations were used to calculate fit statistics. For tests of extension to 100 pg. data, SmartSeq samples were excluded because these samples were not generated using 100 pg. input RNA for cDNA generation and amplification, but by pooling ten 10 pg. diluted replicates before sequencing library preparation, and so were not appropriate for the modeled process. In all cases, validation statistics were calculated based on predictions for genes within covariate ranges used in model fitting and excluding 17 identified large influence genes. For ERCC samples, this meant that four transcripts shorter than 300 nt. were excluded. Also, because the ERCC molecules span a 10^6 range while transcriptomes at a single-cell level span ~10^3 range, 2.5%ile trimming based on input molecules means that only 50 out 92 transcripts were used. While the expected number of input molecules is a very good predictor of gene detection, addition of the remaining independent variables improved prediction (Table B.4). All additional independent covariates also contributed significantly to the model. The model was not sensitive to read-length or paired-end status: it fit data truncated *in silico* to 50 base-pair single-end reads well (Table B.4). The model did not fit ERCC or 100 pg. dilution replicates well (Normalized Chi-square goodness-of-fit test $p < 0.05$); however, it still improved prediction accuracy in these cases compared to using the number of input molecules alone for prediction (Table B.4).

When examining the effect of the number of input molecules on the probability of gene detection, the remaining covariates were set to median values (gene length of 1.05 kilobases, GC content of 0.49, average strength of local secondary structure of -24.7 kcal/mol, no internal A-hexamer, sequencing depth of 17.1 million reads). To calculate the effect of increasing sequencing depth on percent genes recovered, gene detection probabilities were calculated for all genes included in regression analysis (using gene-specific covariate values) at each examined depth. The expected number of genes recovered is the sum of detection probabilities over all genes. To calculate a molecular recovery rate for aRNA with respect to poly-adenylated mRNA

molecules, we fit a logistic model with the same functional form using expected number of input molecules calculated from bulk poly-A HBR samples, gene detection data from aRNA HBR 10 pg. dilution replicates, and fixing the depth coefficient to the value estimated in the above analysis.

3.5.9. Sensitivity outliers

These methods relate to Figure 3.2. We calculated the squared deviance residual for each observation as a measure of fit, using the logistic model described above. The sum of squared deviance residuals is equivalent to the likelihood ratio test statistic comparing the saturated model with respect to the fitted model, and the sum of squared deviance for a subset of observations can be considered the contribution of this set of observations to overall model fit. To find method-specific problematic genes, we calculated the average squared deviance residual for each gene over all samples for each method separately. For each method, we classified genes with average squared deviance residual larger than 4 as outliers. We repeated outlier identification for computationally ambiguous genes within the range of covariates used in model fitting (n=28,270).

3.5.10. Coverage

These methods relate to Figure 3.2. Nucleotide-level coverage was calculated for each gene in the R programming environment (R Development Core Team, 2010) and using Bioconductor libraries GenomicRanges and Rsamtools (Aboyoun et al.; Gentleman et al., 2004; Morgan et al.). Coverage was calculated based on uniquely aligned reads only. Only computationally unambiguous genes were used. Additionally, only genes with a single annotated isoform were used in this analysis.

We calculated the observed per nucleotide coverage scaled by the expected coverage as a function of absolute 3' to 5' position within a gene. HBR and UHR dilution replicates were treated together. Replicates were grouped by institution and by input amount. Each gene in each

sample was considered an independent replicate observation of gene coverage. Genes were filtered to include only those observations with an average of at least 2x coverage per nucleotide. Genes were aligned from the 3' end, so that the per nucleotide sample size decreased from 3' to 5', resulting in increased variance in estimates from 3' to 5'. Nucleotide positions were filtered to include only those with at least 25 replicate observations, which means that for some genes 5' data was excluded. For each gene, per nucleotide coverage was normalized so that the expected coverage at each position was 1x. For each nucleotide position, the expected value is equal to the number of observations at that position and the observed value is the sum of normalized observed values at that position across observations. Using this this scheme, each gene of at least length *i* contributes equally to the observed coverage at position *i*, regardless of expression level. The result is positional observed / expected coverage values.

To examine patterns of gene coverage as a function of expression level, genes were grouped genes by average per nucleotide coverage. We calculated the average per nucleotide coverage for each of 100 equally sized bins from 5' to 3', rather than coverage as a function of absolute nucleotide position as above, in order to observe qualitative coverage patterns occurring at the same relative position along gene bodies. For each gene, bin values were normalized to sum to one so that within an expression level category all genes contribute equally. For an experimental group, positional bins were assigned the average normalized coverage across all genes, observed in any sample within the experimental group, that fell within a given expression level category.

3.5.11. Precision

These methods relate to Figure 3.3. We calculated Pearson pairwise correlation coefficient and Kendall tau pairwise rank correlation coefficient across dilution replicates as a measure of similarity across replicates. The Pearson correlation coefficient is sensitive to large-valued outliers, while the Kendall correlation coefficient is robust. In brief, Kendall correlation is

calculated as follows. For each pair of genes the pair is categorized as concordant if the relative ranks of the gene pair are the same for both samples and discordant otherwise. The coefficient reports the fraction of all pairs that are concordant less the fraction that are discordant. For both correlation coefficients, zeros were treated as missing values, such that only genes observed in both members of a pair were included in the calculation.

To characterize measurement precision, we performed least-squares regression of the empirical standard deviation on the empirical mean. We used computationally unambiguous genes to fit these models. Additionally, we included only genes with >95% probability of presence in a diluted replicate, excluded gene detection outliers and trimmed the upper and lower 2.5%ile by mean value for model fitting. Both the average and standard deviation were log-transformed for normality of residuals. After all filtering, at least 1,100 genes were used to fit the model: $\log_{10}(\text{standard deviation}) = a + b * \log_{10}(\text{mean})$. Sample sizes ranged from 1,149-1,269 genes. A separate model was fit for each experimental group. Because 100 pg. experimental groups have small sample sizes (for most, n<=5) and so provide unstable estimates of variance due to missing values, we performed this analysis on 10 pg. groups only. The NuGen HBR 10 pg. sample size is also quite small (n=4) and was excluded.

To characterize biases in experimental variation we selected a subset of genes where empirical standard deviation was not well predicted by the mean, meaning genes with standardized residuals outside 90% confidence interval of predicted value (assuming a T-distribution with n-3 degrees of freedom for standardized residuals), and a set of "typical" genes, where the gene variance is well predicted by the mean. Typical genes were defined as possessing standardized residuals inside an 80% confidence interval of predicted value. For enrichment tests of GC-content, length, and secondary structure, we calculated the Hodges-Lehmann estimate of difference in location to provide an estimate of the magnitude of in location

between test and background gene. This metric estimates the median difference between the two groups.

We identified outliers with unexpectedly high or low experimental variation as genes with 99.3% confidence interval of predicted value. We considered computationally unambiguous genes, and also extended the analysis to computationally ambiguous genes, excluding those with mean expression outside the range used in model fitting.

Principal components analysis was performed on sample covariance matrix calculated using zero-corrected log-transformed read counts for computationally unambiguous genes with non-zero counts in at least on sample and using the R prcomp function. Each PCA included the appropriate bulk HBR or UHR. RSEM-estimated relative frequencies were normalized to the same scale as the diluted replicates using the DESeq method for estimating size factors, as describe above. Bivariate normal 95% confidence ellipses were calculated for each experimental group using the R dataEllipse function from the car package (Fox and Weisberg, 2011).

3.5.12. Accuracy

These methods relate to Figure 3.4. Sample sizes (number of genes) for analysis in Figure 3.4d-n, given filtering described in plot legend, were the following: HBR: n=1,339 (10 pg.) and 2,797 (100 pg.); UHR: n=1,243 (10 pg.) and 2,614 (100 pg.) As stated, in evaluation of gene measurements in individual dilution measurements genes with zero read counts were excluded. For evaluation of average gene measurements, zero values in individual replicates were retained. RSEM-estimated relative frequencies were treated as true relative expression values for each gene. These were normalized to the same scale as the diluted replicates using the DESeq method for estimating size factors, as describe above. Wide boxes in boxplots of fold deviation in Figure A.1d-f include values for all samples in an experimental group.

To identify method-specific biases in accuracy, we calculated the median fold deviation for each gene across dilution replicates within each experimental group. Genes with fewer than three observations were removed. Of the remaining genes, those with median fold deviation in the upper or lower 5%ile were categorized as overestimated and underestimated, respectively. Remaining genes were used as background for enrichment tests for enrichment. For each method, genes within the upper or lower 1%ile were classified as outlier genes with poor accuracy. Outliers were identified for each experimental group, and then merged across input amounts for each RNA source by taking the union of identified outliers. We repeated outlier identification using computationally ambiguous genes, following the same filtering criteria described above.

To generate density scatter plots of gene read counts in individual dilution replicates, measurements from all 10 pg. dilution replicates for a given method were pooled. The density scatter plots were generated using the R densCols and KernSmooth::bkde2D functions. These functions estimate local density using a binned approximation to a 2 dimensional kernel density with a bivariate Gaussian kernel. $\log_{10}$ read counts were used. For density scatter plots of average read counts, averages were taken separately for HBR and UHR 10 pg. dilution replicates. Averages for HBR and UHR were pooled before density calculation.

3.5.13. Protocol variations

These methods relate to Table 3.2. To evaluate the effect of removing cleaning of initial cDNA, 12 additional HBR 10 pg. dilution replicates were generated. 6 were generated using the same cDNA protocol as the primary aRNA samples, in which first-strand cDNA is cleaned using a MinElute column. 6 were generated without this cleaning step, with adjusted molarity for second-strand cDNA synthesis to accommodate the change in reaction volume. Each set of 6 included 3 replicates generated using 13 rounds of PCR amplification during sequencing library preparation and 3 using 15. In this analysis, differences in PCR treatment were ignored.

To evaluate the effect of reducing rounds of cDNA amplification, 5 additional HBR 10 pg. dilution replicates were generated using 2 rounds of IVT amplification (rather than 3). All amplified material was used as input for sequencing library preparation. Additionally, these samples were generated without initial cDNA purification and using 15 rounds of PCR during sequencing library preparation (rather than 13). These data were compared to 3 replicates generated using 3 rounds of aRNA amplification, and otherwise following the same protocol. To evaluate an optimized aRNA protocol, excluding cleaning and reducing rounds of amplification, the same 5 HBR 10 pg. dilution replicates used to examine the effect of reducing rounds of IVT amplification were compared to the primary HBR 10 pg. aRNA data.

To examine the effect of ERCC addition, 10 replicates beginning with 10 pg. HBR total RNA were amplified using aRNA. In 5, ERCC spike-in controls were added with reference RNA at a final dilution of 1:4,000,000. Samples generated in ERCC optimization showed evidence of cross-contamination, with counts assigned to ERCC transcripts (total ERCC counts: 892-1,457) at appropriate relative abundances for samples generated without addition of ERCC controls.

The effect of strand-specific sequencing was evaluated by re-quantifying aRNA HBR 10 pg. samples using strand information.

CHAPTER 4: Conclusion

4.1. Introduction

In this dissertation we characterized gene expression for single cells from five mammalian cell types using single-cell RNA sequencing data. Additionally, we characterized single-cell RNA sequencing measurements through analysis of data generated in dilution control experiments. In this chapter, we review the key results of this dissertation and close with a discussion of future research directions.

4.2. Single-cell RNA sequencing assessment

We evaluated single-cell RNA sequencing measurements generated by three methods, antisense RNA (aRNA), a customized SmartSeq protocol and a customized NuGen protocol. To do this, we analyzed data generated in two separate dilution experiments. Results in Chapter 2 were based on replicate dilutions of mouse cardiomyocyte RNA to single cell levels (10, 50 and 100 pg.) that were amplified using aRNA and sequenced (Table 2.1 and Figure A.1c-e). In Chapter 3, we considered replicate dilutions of reference RNA (10 and 100 pg) amplified by one of three single-cell RNA sequencing methods (Figure 3.1a). Replicate sequencing data were compared to bulk RNA sequencing measurements of reference RNA. Based on analysis of these data, we provided an overall assessment of the three methods, described measurement biases, characterized measurement reliability as a function of gene expression level, examined the effect of sequencing depth on gene detection, and presented several protocol optimizations.

4.2.1. Overall assessment of measurements

In Chapter 3, the analyzed dilution data demonstrated 89.3 ± 10.6% alignment on average (± Sd.). The distribution of reads across genomic features differed for the three methods (aRNA, SmartSeq and NuGen), with strong enrichment of exonic regions in aRNA

measurements. All methods detected greater than 70% of the expected number of input genes (Figure 3.2a). Based on logistic regression analysis of gene detection data, we found that methods had a 50% probability of detection for genes with an abundance of at least 2 to 4 input molecules (Figure 3.2b, Table B.5). Methods differed in per nucleotide coverage, with NuGen demonstrating the greatest uniformity (Figure 3.2i). Pairwise correlations across diluted replicates were comparable for all methods (Figure 3.3a) and all methods demonstrate similar good pairwise correlation with the reference (Figure 3.4a). In both cases, correlations were higher for larger input amounts. Methods demonstrate similar accuracy at a gene level (Figure 3.4d-f). Average measurements across 10 pg. dilution replicates demonstrated improved median accuracy compared to individual measurements (Figure 3.4d-f).

4.2.2. Biases

In Chapter 3, we evaluated the influence of molecular characteristics on single-cell RNA sequencing measurements. All methods demonstrated qualitatively similar biases in gene detection and quantification, although the degree of bias differed across methods in a manner consistent with differences in molecular protocol (Figure 3.2d-e, Figure 3.4g-j). aRNA exclusively uses poly-T priming for initial RNA capture, and demonstrated the greatest dependence of gene detection on the presence of a poly-A sequence internal to a gene body; however, in general aRNA demonstrated more limited bias in gene detection and quantification than the other methods. NuGen exclusively uses random priming and demonstrated the strongest influence of length on gene detection. SmartSeq uses PCR amplification and demonstrated the largest dependence on GC content. We also identified sets of outlier genes for each method, which demonstrated unreliable measurements in terms of sensitivity, precision or accuracy. In each category, the majority of identified outliers[3] were method-specific.

[3] For computationally unambiguous genes (see section 3.5.3). For this subset of genes, we expect differences to be largely experimental and not do to bioinformatics artifacts.

4.2.3. Measurement reliability and expression level

The reliability of single-cell RNA sequencing expression measurements, as assessed by a diversity of metrics, increased with average expression level. With increasing expression level, the probability of missing values decreases (Figure A.2a, Figure 3.2b), signal-to-noise increases (Figure A.2b), accuracy improves (Figure 3.4k-m), and the distribution of experimental variation approaches normality (Figure A.2c). In Chapter 2, we selected an expression-level threshold to filter unreliable measurements before performing quantitative analyses, beyond which missing values occurred rarely and experimental variation was approximately normal (see Figure A.2a-d). We conservatively estimated that aRNA measurements were reliable for genes expressed at a level of at 4–9 input molecules[4]. In Chapter 3, we showed the functional relationship between input molecules and read counts for each of the amplification methods (Figure 3.4k-m) and found that this relationship was roughly linear for expression levels greater than ~5–10 expected input molecules, concordant with Chapter 2 estimates. This corresponds to ~4,000–8,000 reliably measured genes.

4.2.4. Depth of sequencing and sensitivity

Despite the small number of input molecules in a 10 pg. dilution replicate, in Chapter 3 we found that sequencing depth had a significant effect on gene detection, with the odds of gene detection increasing substantially with increasing sequencing depth until a depth of ~15 million reads (Figure 3.2)[5]. For these dilution data, increasing sequencing depth from 10 to 15 million reads resulted in ~25% expected gain in detected genes. Gains in detection at this high level of coverage (~50x) are surprising, and we speculate that this effect may be related to the greater

[4] Assuming 150,000 total input molecules.

[5] In Chapter 2, we estimated that little sensitivity was gained beyond 5 million uniquely assigned exonic reads (Figure A.1a), which corresponds to a total sequencing depth of ~16 million.

dispersion of gene abundance estimates observed in the amplified replicates compared to the reference distribution (Figure 3.4k-m).

4.2.5. Protocol optimizations

Our analysis of aRNA optimizations in Chapter 3 demonstrated that reducing cleaning steps, rounds of amplification, and the extent of dilution on library preparation lead to overall measurement improvement, and suggest that these strategies may be productive avenues for methods development (Table 3.2).

4.3. Survey of transcriptome heterogeneity across single cells from diverse mammalian tissues

In Chapter 2, we used single-cell RNA sequencing data to characterize single-cell transcriptome heterogeneity in five mammalian cell types: brown adipocytes, cardiomycotes, cortical and hippocampal pyramidal neurons, and serotonergic neurons. We found that single-cell transcriptome complexity and patterns of gene expression variation differed across cell types. Our results suggested that ubiquitous expression across cells may be indicative of critical gene function, and that expression variation in some cases may reflect dynamic cell behaviors. Several of our results provide evidence suggesting that the extent of expression variation may be under regulatory control. We observed broad correlation of gene expression variation in mouse pyramidal neurons with that in rat, consistent with the hypothesis that expression variation may be conserved.

4.3.1. Tissue-specific transcriptome characteristics

Cell types differed in global transcriptome properties, such as the total number of genes expressed and the fraction of the transcriptome that is non-genic. Using these metrics, we found that pyramidal neurons demonstrated unique transcriptional complexity compared to other cell types (Figure 2.1c-d). Because ubiquitous expression across cells may provide an indicator of

phenotypic importance, we identified genes observed consistently highly expressed in all cells of a given type, as well as those universally expressed. Genes consistently expressed at a high expression level within a cell type were enriched for cell type-specific function, while genes with universal expression were enriched in housekeeping function (Figure 2.2a). Genes observed in all cells were significantly more likely than background to result in prenatal lethality on genetic disruption, while those highly expressed within a cell type were significantly more likely than background to produce tissue-specific defects, with the exception of cardiomyocytes (Figure 2.2b).

4.3.2. Patterns of single-cell expression variation

As a measure of single-cell expression heterogeneity, we calculated a studentized F-statistic of biological variation over experimental variation observed at matched expression strength, to account for the expression dependence of measurement noise. Every cell type contained a subset of highly variable genes by this measure, with an F-statistic greater than 10 (Figure 2.3a). As described in Chapter 1, large expression variability within a cell type may characterize genes critical for cell-type specific dynamic phenotypes (see section 1.3.2). Indeed, when we examined the 5% most variable genes for each cell type, we found that functional categories relevant for plastic phenotypes were enriched in a cell-type specific manner, including categories such as tissue development, cell division and cell-to-cell signaling (Figure 2.3b). Conversely, when we examined the relative variability of genes grouped by broad functional category, we found that transcription factors and ion channels, important for dynamic behavior and response to the environment, demonstrated greater variability than ribosomal and metabolic genes (Table 2.2).

4.3.3. Evidence suggesting control over the extent of expression variation

If patterns of gene expression variation among individual cells have functional relevance, these patterns may be controlled (as discussed in section 1.5). We identified a subset of genes to be

highly variable in one cell type while demonstrating uniform expression in a second cell type, suggesting underlying differences in regulatory dynamics and potential differences in gene function (Figure 2.3e). This set includes individual cases suggestive of critical cell type-specific function. One regulatory mechanism that could modulate expression variation is the rate of transcript decay. Using publicly available data, we found significant association of RNA decay rate with the outlier-sum statistic (a metric to identify extreme variability, Figure 2.3f). If the extent of single-cell expression variation is important for tissue function, levels of variation may be conserved across species. The sets of gene identified as ubiquitously highly expressed in rat and in mouse demonstrated significant similarity (Figure 2.2c-d), and F-statistic values calculated in rat and in mouse pyramidal neurons were significantly associated (-d). Sample sizes for comparative analysis were small, and further studies will be needed to confirm this result; however, collectively and in conjunction with the functional coherence among highly variable genes and the association of consistent genes with tissue-specific mutant phenotypes, these results suggest that the extent of expression variation may be controlled.

4.4. Future outlook

Recent years have seen burgeoning interest in single-cell RNA sequencing, coupled with continuous improvements in experimental and analytic methods. We expect this advancement will continue at a rapid pace, expanding the catalogue of characterized cell types and generating insights into the mechanisms of gene expression; however, the frontier of single-cell research may be in the coupling of single-cell and tissue biology and, in particular, the relationship between tissue behavior and single-cell heterogeneity.

Research in this area faces many obstacles and progress is uncertain. There are limited methods to concurrently measure single-cell expression profiles and cell phenotype, and limited analytic techniques to infer this relationship based on available data. Conceptual work is needed to hypothesize how tissue behavior arises from heterogeneous single cells, drawing on expertise

across biological disciplines and including gene expression and regulation, cell biology, physiology, and ecology and evolution.

Pursuing these hypotheses will require novel experimental design and methodology. One fruitful avenue may be further testing for conservation of tissue-specific expression variation across species. Such conservation would be consistent with expression variation contributing to organismal fitness; however, direct experimental testing will be necessary to establish that a relationship between variation and tissue behavior exists. Conservation analysis could be used to select a subset of promising genes to test experimentally. One experimental test for a relationship between population behaviors and single-cell expression heterogeneity that has proved useful in the literature is to measure relevant cell traits for single cells in isolation, and also in the context of a population, concurrently with expression for one or a few genes (Patil et al., 2015; Shalek et al., 2014; Xue et al., 2015). In these experiments, differences across contexts in the measured cell traits suggests that population behavior arises from the collection of constituent expression states. More conclusive experiments will require the development of methods to manipulate gene expression heterogeneity across a population, perhaps through manipulation of mechanism capable of controlling the extent of variation. In Chapter 1 we reviewed two experiments that may provide prototypes for this methods development, where genetic and epigenetic manipulations were performed that modulated the extent of gene expression variation across cells (see section 1.5).

Though the challenges are formidable, pursuit of the relationship between single-cell gene expression and tissue behavior has the potential to revolutionize our understanding of health and disease in multicellular organisms. Progress on this topic will require a community of researchers and substantial time. It is our hope that the work in this dissertation may contribute to the early stages of this research: to improvement in the analysis and interpretation of single-cell RNA sequencing data, to progress in the association of multi-genic expression variation and cell

phenotype, and to the conceptual discussion around how tissue behavior stems from the behavior of heterogeneous single cells.

APPENDIX A: Supplemental material for Chapter 2

A.1. Supplemental tables

Additional tables are contained in supplemental file 1. Legends are below.

Table A.1 RNA sequencing and quality control data for each single-cell sample

See section 2.4.4 for details about quality indicators used.

Table A.2 Summary of single-cell transcriptome characteristics

See section 2.2 for details on cell size adjustment.

Table A.3 Identified gene sets

Includes lists of private genes (genes observed in only one single-cell sample); universally common genes (genes observed in every single-cell sample); cell-type specific common (genes abundantly expressed in every single-cell sample of a specific cell type, excluding universally common genes); 5% most variable genes as identified by F-statistic; most variable genes by outlier sum statistic; most normally expressed genes by Shapiro–Wilk statistic; and differentially variable genes across cell types. Lists are based on mouse data.

Table A.4 Association between common expression and mutant phenotypes

Contingency tables show counts of genes with common expression and for remaining expressed genes (with at least one read in at least one sample for indicated sample set), categorized as causing indicated mutant phenotype or not. Tables are shown for all significant pairwise associations (Chi-square tests of independence with Bonferonni-corrected p-values < 0.05) and also for cardiomyocyte samples and mutant phenotypes. Phenotypic annotations were accessed from the Mammalian Phenotype Ontology on June 23, 2013 (Eppig et al., 2012; Smith and Eppig, 2009). Annotation IDs assigned to each mutant phenotypic category can be found in Table A.7.

Table A.5 Within- and between-cell type transcriptome variance

Computed on variance-stabilized frequencies. Only genes with an average relative frequency of greater than 6.3×10^{-5} across all cells are included. Variance stabilization: asin(2x-1) - asin(-1).

Table A.6 Enrichment of Gene Ontology (GO) categories among genes identified by expression pattern

Genes identified by the F-statistic, variable genes identified by the outlier-sum-statistic, and normally expressed genes by cell type.

Table A.7 Mammalian Phenotype Ontology annotations used in mutation analysis

Annotations considered indicative of prenatal lethality or of tissue-specific mutant phenotypes (Smith and Eppig, 2009).

A.2. Supplemental figures

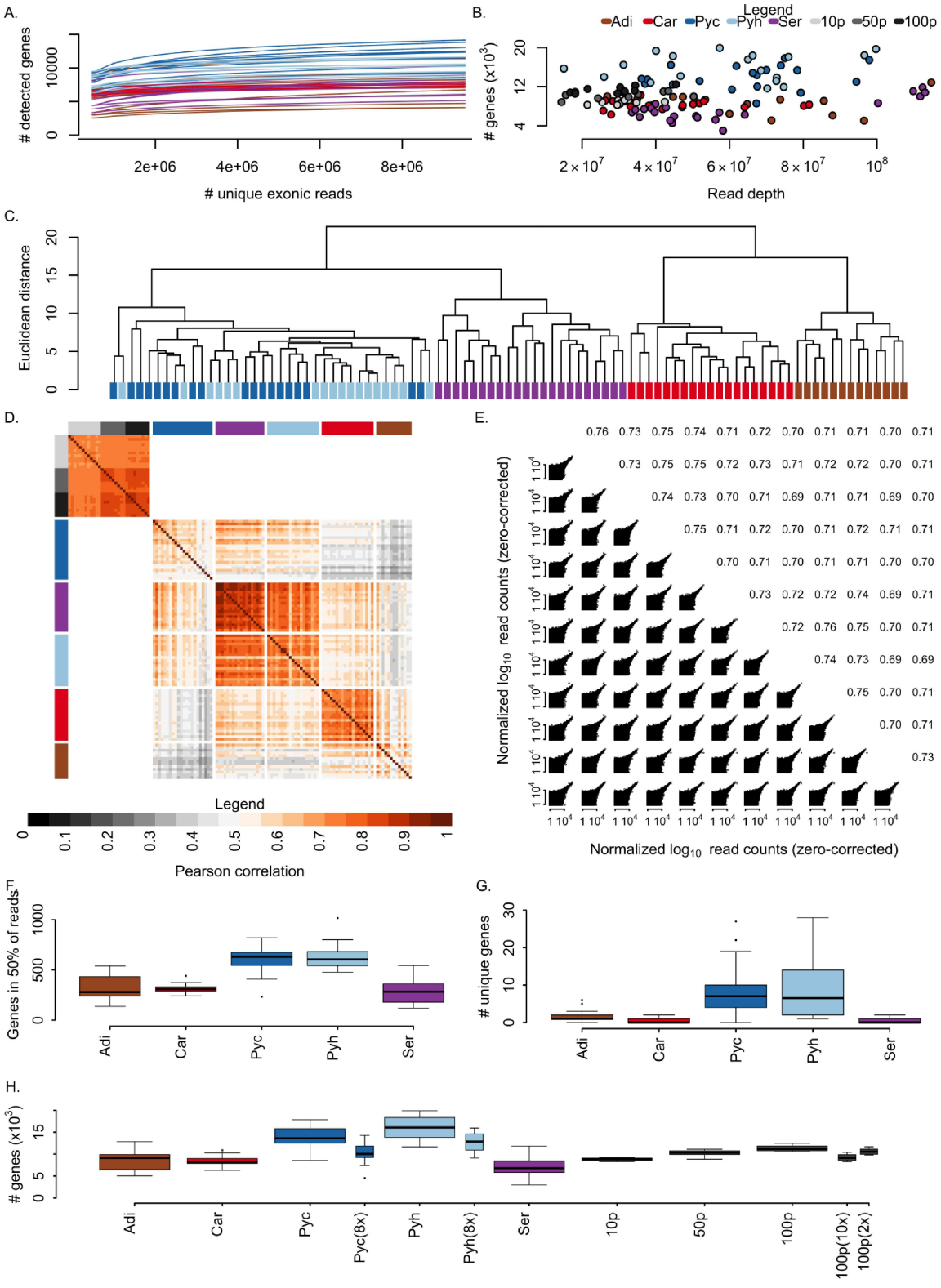

A.
detected genes
0
10000
2e+06
4e+06
6e+06
8e+06
unique exonic reads
B.
Legend
Adi Car Pyc Pyh Ser 10p 50p 100p
genes (x10^3)
4
12
20
2 × 10^7
4 × 10^7
6 × 10^7
8 × 10^7
10^8
Read depth
C.
Euclidean distance
0
5
10
15
20
D.
Legend
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Pearson correlation
E.
0.76 0.73 0.75 0.74 0.71 0.72 0.70 0.71 0.71 0.70 0.71
0.73 0.75 0.75 0.72 0.73 0.71 0.72 0.72 0.70 0.71
0.74 0.73 0.70 0.71 0.69 0.71 0.71 0.69 0.70
0.75 0.71 0.72 0.70 0.71 0.72 0.71 0.71
0.70 0.71 0.70 0.71 0.71 0.70 0.70
0.73 0.72 0.72 0.74 0.69 0.71
0.72 0.76 0.75 0.70 0.71
0.74 0.73 0.69 0.69
0.75 0.70 0.71
0.70 0.71
0.73
1 10^4
Normalized log10 read counts (zero-corrected)
F.
Genes in 50% of reads
0
500
1000
Adi
Car
Pyc
Pyh
Ser
G.
unique genes
0
10
20
30
Adi
Car
Pyc
Pyh
Ser
H.
genes (x10^3)
0
5
15
Adi
Car
Pyc
Pyc(8x)
Pyh
Pyh(8x)
Ser
10p
50p
100p
100p(10x)
100p(2x)

Figure A.1 Dataset quality and transcriptome characteristics.

a-b Single cell samples were sequenced to a sufficient depth. (a) Pseudo-single-cell RNA sequencing libraries with a range of sequencing depths, generated by randomly subsampling reads from single-cell RNA sequencing libraries. 100 pseudo-libraries were generated for nine cells of each mouse cell type ranging from 0.5 million to 9.5 million uniquely aligning exonic reads. (b) Number of observed genes as a function of sequencing depth for single cell dataset. **c** Complete-linkage hierarchical clustering of single-cell samples based on Euclidean distance between gene expression profiles using $\log_{10}$ zero-corrected normalized read counts for marker gene expression. **d-e**. Pairwise correlations of zero-corrected normalized read counts for all detected genes (greater than zero reads in any sample) on a $\log_{10}$ scale. (d) Pairwise Pearson correlation coefficients for entire mouse dataset. Sample tissue or dilution input amount for a given row is indicated by the colored bar to the left of the heatmap and for a given column by the colored bar above the correlation heatmap. (e) Scatter plots of technical amplification replicates beginning with 10 picograms (pg) of total RNA. Upper triangle contains Pearson correlation coefficient. **f-g** Tissues differ in single cell transcriptome characteristics. (f) The number of highly expressed genes comprising 50% of reads by cell type. (g) The number of genes found only in a single cell by cell type. **h** The number of genes detected by cell type after correction for cell size. An adjusted detection threshold has been applied to each pyramidal neuron, removing all genes with expression below 8 times the minimum observed relative frequency. 100 picogram dilution replicates have been similarly corrected for a 10-fold and 2-fold difference in input RNA. Sample sizes, colors and abbreviations: brown adipocytes (n=13, *brown*, *Adi*); cardiomyocytes (n=19, *red*, *Car*); pyramidal neurons, cortex (n=19, *dark blue*, *Pyc*); pyramidal neurons, hippocampus (n=18, *light blue*, *Pyh*); serotonergic neurons, dorsal raphe (n=22, *purple*, *Ser*); dilution controls starting with 100 pg (n=9, *dark gray*, *100p*), 50 pg (n=9, *gray*, *50p*), or 10 pg total RNA (n=12, *light gray*, *10p*).

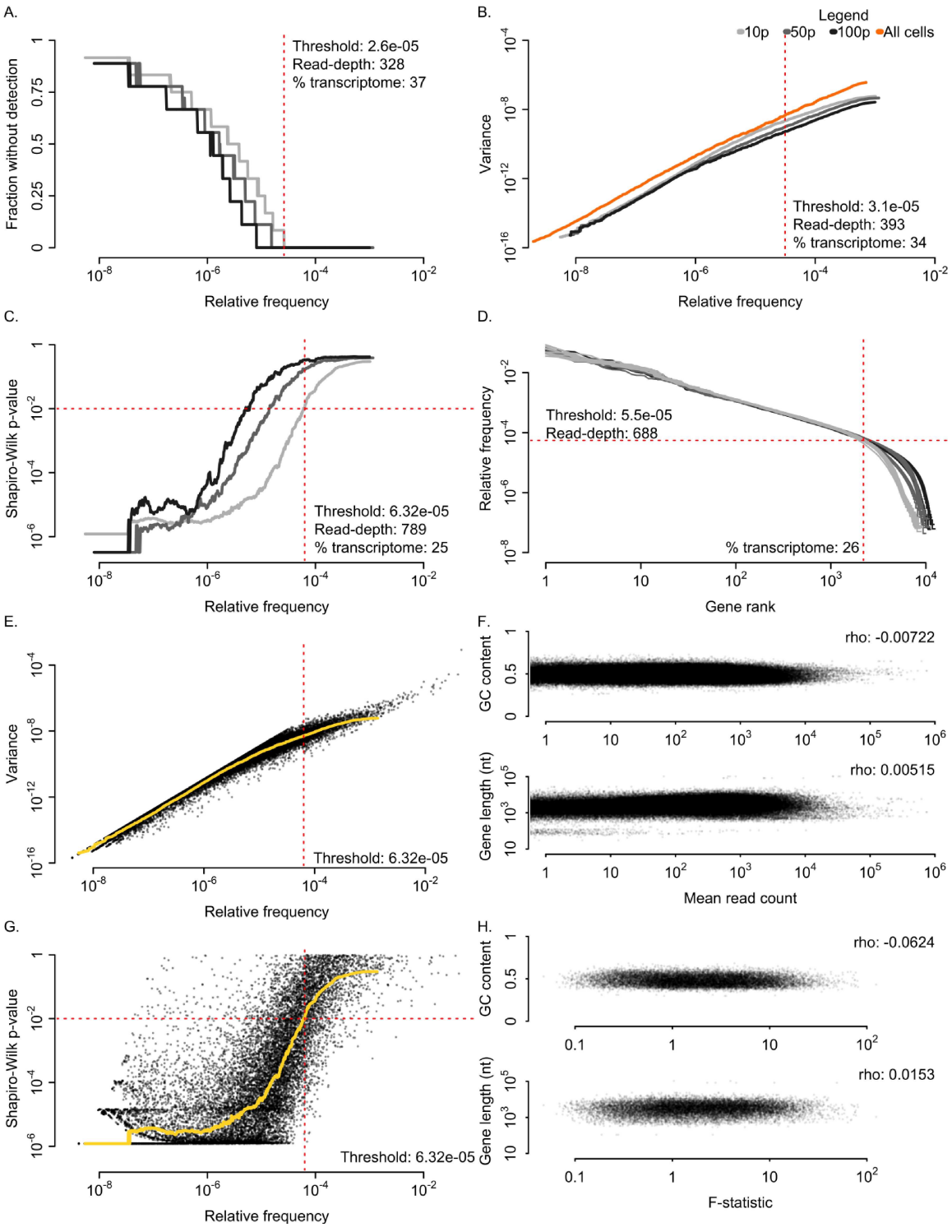

A.
Fraction without detection
Threshold: 2.6e-05
Read-depth: 328
% transcriptome: 37
Relative frequency
B.
Legend
10p
50p
100p
All cells
Variance
Threshold: 3.1e-05
Read-depth: 393
% transcriptome: 34
Relative frequency
C.
Shapiro-Wilk p-value
Threshold: 6.32e-05
Read-depth: 789
% transcriptome: 25
Relative frequency
D.
Relative frequency
Threshold: 5.5e-05
Read-depth: 688
% transcriptome: 26
Gene rank
E.
Variance
Threshold: 6.32e-05
Relative frequency
F.
GC content
rho: -0.00722
Gene length (nt)
rho: 0.00515
Mean read count
G.
Shapiro-Wilk p-value
Threshold: 6.32e-05
Relative frequency
H.
GC content
rho: -0.0624
Gene length (nt)
rho: 0.0153
F-statistic

Figure A.2 Accounting for technical characteristics of aRNA sequencing.

a-d Selection of an expression level threshold for reliable quantification. Reported threshold read-depths are average values across all biological samples. '% transcriptome' indicates the percent of the expressed transcriptome with expression levels greater than threshold, averaged across biological samples. In panels (a-c), solid lines indicate median value across 500 neighbors by expression level for each group. (a) Fraction of replicates without detection of gene expression as a function of expression level. Vertical dotted line indicates threshold beyond which genes are reliably detected across replicates. (b) Variance across biological and dilution replicates as a function of expression level. Vertical dotted line indicates threshold beyond which genes' median variation across biological replicates is at least twice the median variation observed across 10 picogram (pg.) dilution replicates. (c) Normality of experimental variation as a function of expression level and amount of input RNA. Vertical dotted line indicates threshold beyond which Shapiro-Wilk p-value > 0.01 consistently across 10 pg. dilution replicates. (d) Expression level (relative frequency) vs. gene expression rank. Horizontal dotted line indicates estimated threshold beyond which frequency decreases more rapidly with gene rank in the 10 pg. dilution controls. Solid lines represent individual samples. (e) Variation as a function of expression level across 10 pg. dilution controls. Yellow line indicates median variation of 500 neighboring genes by expression level, which is used as an estimate of experimental variation as a function of expression level. (f) Scatter plots of gene traits and gene expression measurements. Mean read counts were calculated separately for each cell type and for dilution controls. (g) Experimental variation normality as a function of expression level for 10 pg. dilution controls. Vertical dotted line indicates threshold beyond which Shapiro-Wilk p-value > 0.01 consistently across 10 pg. dilution replicates. (h) F-statistic (ratio of single-cell variation to experimental variation) does not depend on gene GC content or gene length. Scatter plots of gene traits and F-statistic values. Mean F-statistic values were calculated separately for each cell type. Sample sizes, colors and abbreviations: 10 pg. dilution replicates (n=12, *light gray*, *10p*); 50 pg. replicates (n=9, *gray*, *50p*); 100 pg. replicates (n=9, *dark gray*, *100p*); all single cell samples (n=91, *orange*, *All cells*).

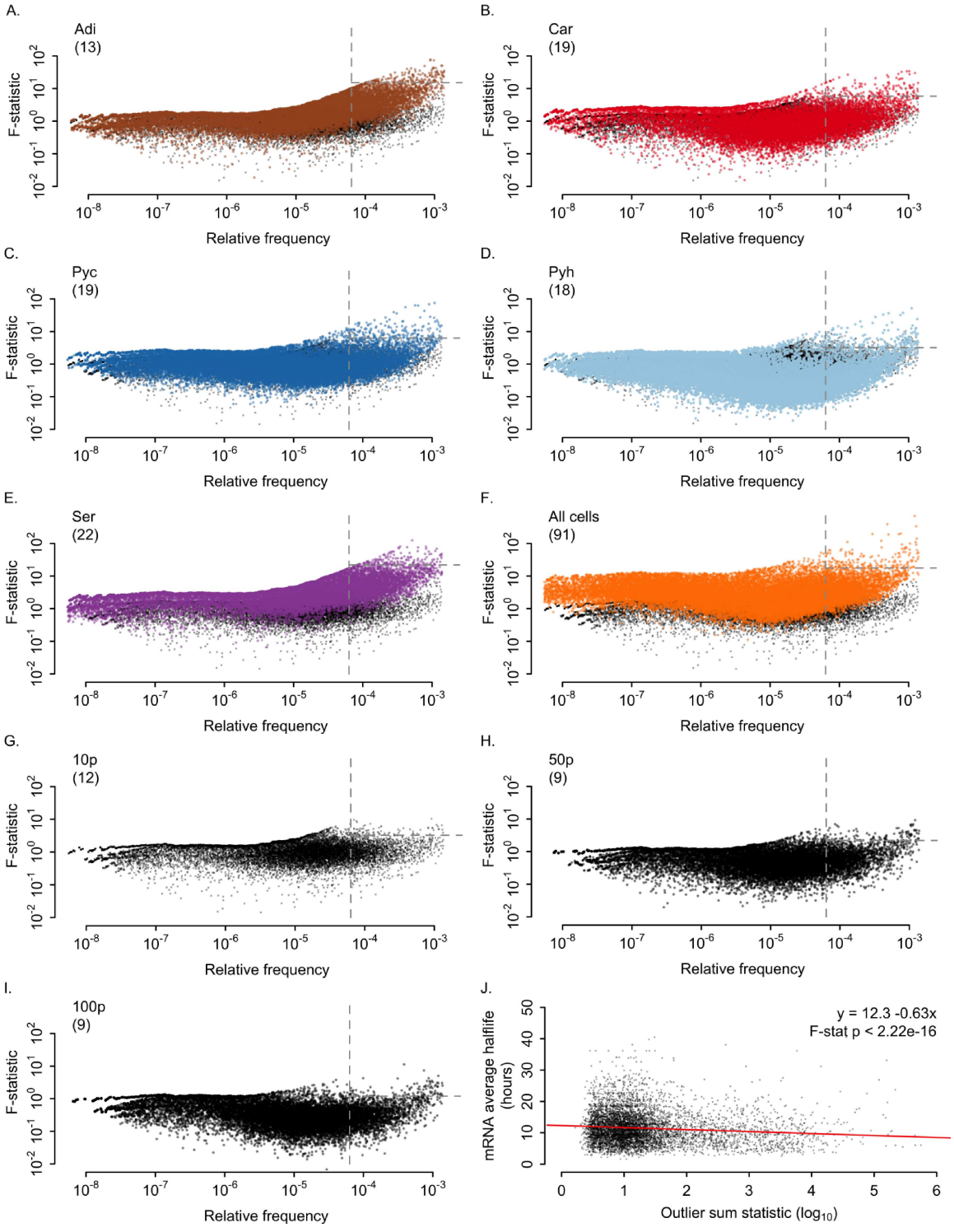

A.
Adi
(13)
B.
Car
(19)
C.
Pyc
(19)
D.
Pyh
(18)
E.
Ser
(22)
F.
All cells
(91)
G.
10p
(12)
H.
50p
(9)
I.
100p
(9)
F-statistic
Relative frequency
J.
y = 12.3 -0.63x
F-stat p < 2.22e-16
mRNA average halflife (hours)
Outlier sum statistic (log10)

Figure A.3 A subset of genes demonstrates variable expression in each examined tissue type. **a-i** Scatter plots of F-statistic as a function of gene expression level for each experimental group. Vertical hashed line indicates quality control threshold used for variation analysis. Genes with relative frequencies below this threshold were not included in variation analysis. Horizontal hashed line indicates the top 5% of included genes by F-statistic value for each cell type or dilution control. The F-statistic across 10 pg. dilution controls is shown in black in panels (a-f). **j** Scatter plot of average mRNA half-life in hours and outlier-sum statistic values across all single-cell samples. Sample sizes, colors and abbreviations: brown adipocytes (n=13, *brown*, *Adi*); cardiomyocytes (n=19, *red*, *Car*); pyramidal neurons, cortex (n=19, *dark blue*, *Pyc*); pyramidal neurons, hippocampus (n=18, *light blue*, *Pyh*); serotonergic neurons, dorsal raphe (n=22, *purple*, *Ser*); all single cell samples (n=91, *orange*, *All cells*); 10 pg. dilution replicates (n=12, *black*, *10p*); 50 pg. dilution replicates (n=9, *black*, *50p*); 100 pg. dilution replicates (n=9, *black*, *100p*).

APPENDIX B: Supplemental material for Chapter 3

B.1. Supplemental tables

Additional tables are contained in supplemental file 2. Legends are below.

Table B.1 Control dataset sample identification, protocol information, and RNA sequencing stats

Experimental group, protocol information and RNA sequencing statistics for each sample used in primary analyses. Alignment statistics were based on STAR alignment to hg19 and were with respect to reads retained after trimming for primer or poly-A sequences (Dobin et al., 2012; Fisher and Kim, 2015). Material supports section 3.3.1.

Table B.2 Optimization dataset sample identification, protocol information, and RNA sequencing stats

As table B.1 for samples used in protocol optimization analyses. Material supports section 3.3.6.

Table B.3 Gene detection logistic regression model

See model details in section 3.5.8. *Abbreviations*: *M* expected number of input molecules; *L* gene length (kilobases); *G* gene GC content; *S* strength of gene local secondary structure (kilocalories per mole); *hasA* presence of A-hexamer internal to gene body; *D* Depth (per 10,000,000 reads); *S.E.* standard error; *Wald Z* Wald test statistic; *Pr(>|Z|)* Wald test p-value. Material supports Figure 3.2.

Table B.4 Gene detection logistic regression fit and validation

Model was fit using randomly selected 90% of 10 pg. data, excluding 17 large influence genes. Fit was evaluated on the remaining 10% of the data. Fit was also evaluated on sequence data that was *in silico* truncated to 50 base pair single end ("Truncated"), ERCC read counts ("ERCC"), and 100 pg. dilution replicates ("100 pg."). AUC (Area under receiver operating characteristic curve) reported as mean values ± 2 Sd. calculated over 10,000 bootstrap samples. AUC (molecules) predicts detection based on number of input molecules alone. See section 3.5.8 for further details. Material supports Figure 3.2.

Table B.5 Probability of gene recovery

Based on model described in section 3.5.8. Remaining covariates set to median value. Material supports Figure 3.2.

Table B.6 Gene detection outliers

Genes that are problematic for detection. See section 3.5.9 for classification of outliers. "Gene set" indicates whether gene is classified as computationally unambiguous (1) or not (2). "Detected / undetected" indicates whether the gene is unexpectedly observed (D) or unexpectedly unobserved (U). Material supports Figure 3.2.

Table B.7 Precision outliers

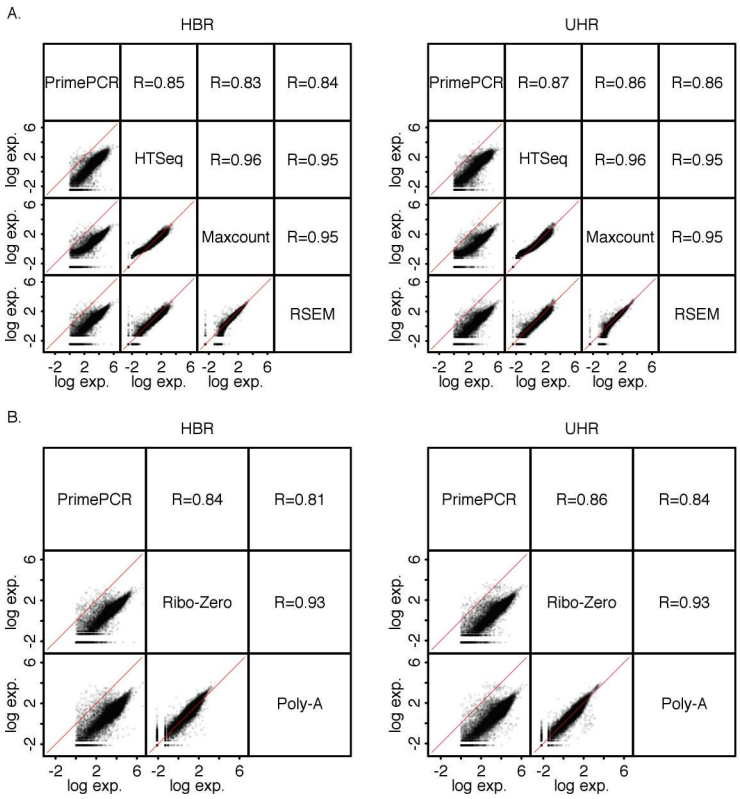
Genes with standardized residual outside a 99.3% confidence interval, with respect to regression of standard deviation on the mean (see section 3.5.11). "Gene set" indicates whether gene is classified as computationally unambiguous (1) or not (2). Only genes whose mean is within the range of fitted model were included. Column values indicate whether indicate whether the gene standard deviation is unexpectedly low (L) or high (H), given mean. Material supports Figure 3.3.

Table B.8 Accuracy outliers

Genes were identified as accuracy outliers if its median fold deviation, taken across dilution replicates, was contained in the upper or lower 1%ile of all considered genes (see section 3.5.12). Columns labeled by single-cell protocol contain an "H" if a gene was identified as an overestimated outlier, and an “L” if a gene was identified as an underestimated outlier. "Gene set" indicates whether gene is classified as computationally unambiguous (1) or not (2). Material supports Figure 3.4.

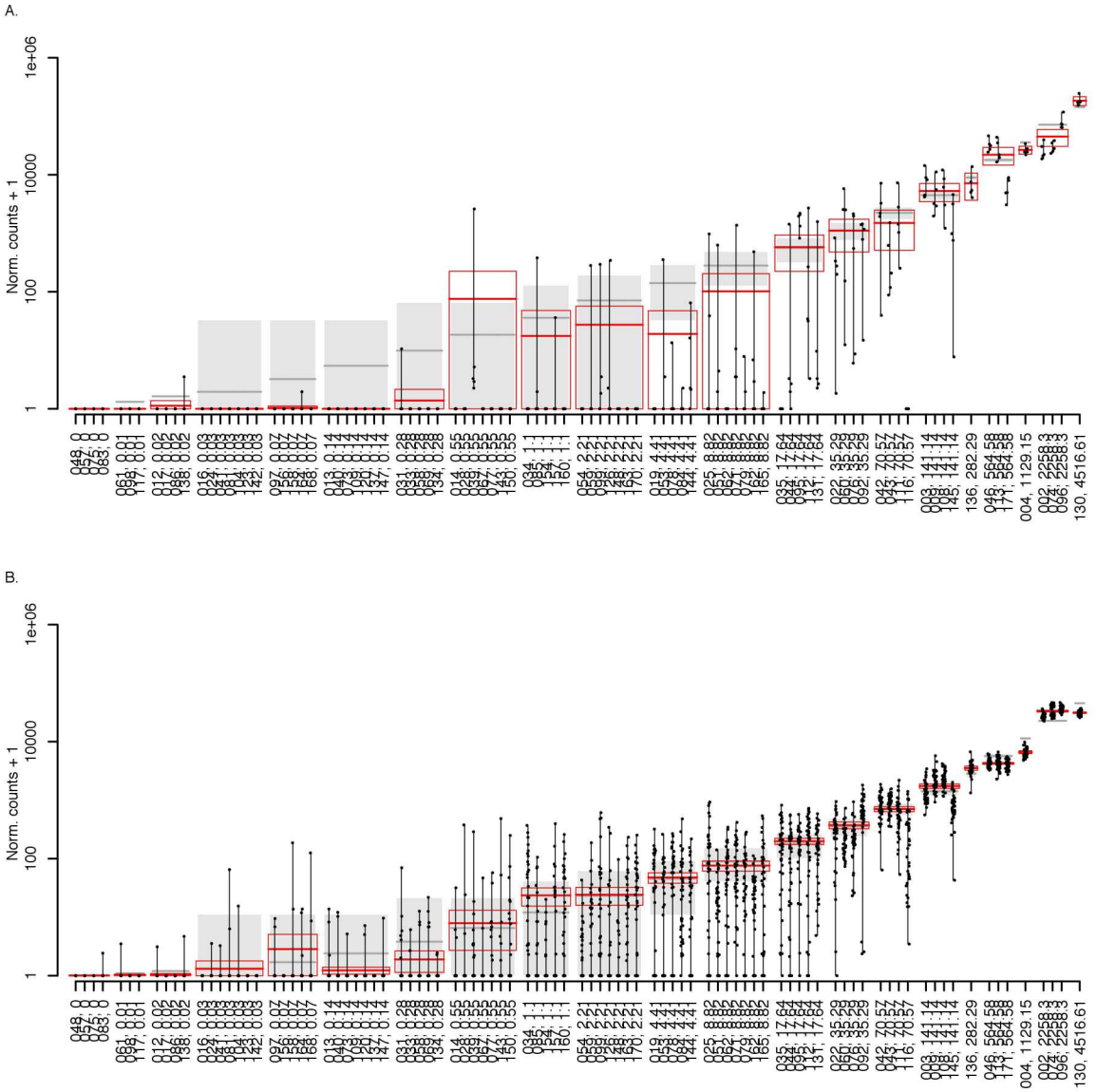
B.2. Supplemental figures

Figure B.1 Accuracy and robustness of estimated reference HBR and UHR RNA expression levels



A. Consistency of abundance estimates by three quantification algorithms relative to publicly available PrimePCR measurements (see 3.5.1). Scatters show $\log_{10}$ reads per million (HTSeq (Anders, 2010) and Maxcounts (Finotello et al., 2014)), $\log_{10}$ transcripts per million (*RSEM*), or $\log_{10}$ molecules (*PrimePCR*). Upper quadrants indicate Pearson correlation (*R*) of log-transformed estimates. Pairwise zeros were treated as missing values. Estimates were based on combined raw reads from 3 bulk reference samples generated using ribosomal depletion for each HBR and UHR. RSEM estimates were used as reference throughout. **B.** Accuracy and robustness of expression estimates across library preparation methods: ribosomal-depletion (combined n=3 samples per HBR and UHR) and poly-A RNA selection (combined n=4 samples per source). See section 3.5.1 for sample information. Scatters as in A using RSEM expression level estimates for each library preparation method. ribosomal-depletion samples were used as reference throughout. Abbreviations: Human Brain Reference (*HBR*), Universal Human Reference RNA (*UHR*). Material supports section 3.3.2.

Figure B.2 Normalized read counts and expectation for ERCC transcripts



A-B. ERCC transcripts are found along the x-axis, ordered by expected number of input molecules. Axis labels are in the format of "ERCC spike-in ID, expected number of input molecules". Points indicate the normalized read count for one transcript in one sample. Horizontal gray lines and background gray boxes indicate the expected normalized read count and a 95% CI under a Poisson model of dilution. Wide red horizontal lines indicate mean normalized read counts across all ERCC transcripts with a common expected number of input molecule, and red boxes indicate mean ± 2 x s.e.m. (A) aRNA. (B) SmartSeq. Material supports section 3.3.6.

BIBLIOGRAPHY

Aboyoun, P., Pages, H., and Lawrence, M. GenomicRanges: Representation and manipulation of genomic intervals.

Acar, M., Mettetal, J.T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. Nat Genet *40*, 471–475.

Ahrends, R., Ota, A., Kovary, K.M., Kudo, T., Park, B.O., and Teruel, M.N. (2014). Controlling low rates of cell differentiation through noise and ultrahigh feedback. Science *344*, 1384–1389.

Ambler (original), G., and Benner (modified), A. (2015). mfp: Multivariable Fractional Polynomials.

Anders, S. (2010). HTSeq: Analysing high-throughput sequencing data with Python (EMBL Heidelberg (Genome Biology Unit)).

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma. Oxf. Engl. *31*, 166–169.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Bajikar, S.S., Fuchs, C., Roller, A., Theis, F.J., and Janes, K.A. (2014). Parameterizing cell-to-cell regulatory heterogeneities via stochastic transcriptional profiles. Proc. Natl. Acad. Sci. *111*, E626–E635.

Beaumont, H.J.E., Gallie, J., Kost, C., Ferguson, G.C., and Rainey, P.B. (2009). Experimental evolution of bet hedging. Nature *462*, 90–93.

Benayoun, B.A., Pollina, E.A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E.D., Devarajan, K., Daugherty, A.C., Kundaje, A.B., Mancini, E., et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell *158*, 673–688.

Bieler, J., Cannavo, R., Gustafson, K., Gobet, C., Gatfield, D., and Naef, F. (2014). Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Mol. Syst. Biol. *10*.

Brady, G., Barbara, M., and Iscove, N.N. (1990). Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. Methods Mol Cell Biol *2*, 17–25.

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

Briney, B.S., and Jr, J.E.C. (2013). Secondary mechanisms of diversification in the human antibody repertoire. Front. Immunol. *4*.

Buchhalter, J.R., and Dichter, M.A. (1991). Electrophysiological comparison of pyramidal and stellate nonpyramidal neurons in dissociated cell culture of rat hippocampus. Brain Res. Bull. *26*, 333–338.

Buckley, P.T., Lee, M.T., Sul, J.-Y., Miyashiro, K.Y., Bell, T.J., Fisher, S.A., Kim, J., and Eberwine, J. (2011). Cytoplasmic intron sequence-retaining transcripts (CIRTs) can be dendritically targeted via ID element retrotransposons. Neuron *69*, 877–884.

Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., and Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell. Nat. Biotechnol. *33*, 155–160.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., et al. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. *28*, 264–278.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., AmiGO Hub, and Web Presence Working Group (2009). AmiGO: online access to ontology and annotation data. Bioinforma. Oxf. Engl. *25*, 288–289.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*.

Chiu, I.M., Barrett, L.B., Williams, E.K., Strochlic, D.E., Lee, S., Weyer, A.D., Lou, S., Bryman, G., Roberson, D.P., Ghasemlou, N., et al. (2014). Transcriptional profiling at whole population and single cell levels reveals somatosensory neuron molecular diversity. eLife.

Cornelison, D.D., and Wold, B.J. (1997). Single-cell analysis of regulatory gene expression in quiescent and activated mouse skeletal muscle satellite cells. Dev. Biol. *191*, 270–283.

Dar, R.D., Hosmane, N.N., Arkin, M.R., Siliciano, R.F., and Weinberger, L.S. (2014). Screening for noise in gene expression identifies drug synergies. Science *344*, 1392–1396.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

Derrien, T., Estellé, J., Marco Sola, S., Knowles, D.G., Raineri, E., Guigó, R., and Ribeca, P. (2012). Fast Computation and Applications of Genome Mappability. PLoS ONE *7*, e30377.

Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015a). Integrated genome and transcriptome sequencing of the same cell. Nat Biotech *33*, 285–289.

Dey, S.S., Foley, J.E., Limsirichai, P., Schaffer, D.V., and Arkin, A.P. (2015b). Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. Mol. Syst. Biol. *11*.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics bts635.

Dueck, H., Eberwine, J., and Kim, J. (2015a). Variation is Function: Searching for Meaning in Single Cell Variation. Submiss.

Dueck, H., Khaladkar, M., Kim, T.K., Spaethling, J.M., Francis, C., Suresh, S., Fisher, S.A., Seale, P., Beck, S.G., Bartfai, T., et al. (2015b). Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. Genome Biol. *16*.

Dueck, H., Ai, R., Dominguez, R., Evgrafov, O., Fan, J.-B., Fisher, S., Francis, C., Herrnstein, J., Kim, T.K., Kim, H., et al. Assessment of single cell RNA sequencing using reference RNA diluted to single cell levels. *In progress.*

Durruthy-Durruthy, R., Gottlieb, A., Hartman, B.H., Waldhaus, J., Laske, R.D., Altman, R., and Heller, S. (2014). Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. Cell *157*, 964–978.

Eberwine, J., and Crino, P. (2001). Analysis of mRNA populations from single live and fixed cells of the central nervous system. Curr. Protoc. Neurosci.

Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P., and Spector, D.L. (2014). Random Monoallelic Gene Expression Increases upon Embryonic Stem Cell Differentiation. Dev. Cell *28*, 351–365.

Eguchi, J., Yan, Q.-W., Schones, D.E., Kamal, M., Hsu, C.-H., Zhang, M.Q., Crawford, G.E., and Rosen, E.D. (2008). Interferon regulatory factors are transcriptional regulators of adipogenesis. Cell Metab. *7*, 86–94.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. Nature *467*, 167–173.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183–1186.

Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A., Richardson, J.E., and Mouse Genome Database Group (2012). The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res. *40*, D881–D886.

Erwin, J.A., Marchetto, M.C., and Gage, F.H. (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. Nat. Rev. Neurosci. *15*.

Fang, M., Xie, H., Dougan, S.K., Ploegh, H., and van Oudenaarden, A. (2013). Stochastic cytokine expression induces mixed T helper cell States. PLoS Biol. *11*.

Finotello, F., Lavezzo, E., Bianco, L., Barzon, L., Mazzon, P., Fontana, P., Toppo, S., and Di Camillo, B. (2014). Reducing bias in RNA sequencing data: a novel approach to compute counts. BMC Bioinformatics *15*, S7.

Fisher, S., and Kim, J. (2015). ngs_TRIM.

Fox, J. (2003). Effect Displays in R for Generalised Linear Models. J. Stat. Softw. *8*, 1–27.

Fox, J., and Weisberg, S. (2011). An R Companion to Applied Regression (Thousand Oaks CA: Sage).

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Grant, G.R., Farkas, M.H., Pizarro, A.D., Lahens, N.F., Schug, J., Brunk, B.P., Stoeckert, C.J., Hogenesch, J.B., and Pierce, E.A. (2011). Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinforma. Oxf. Engl. *27*, 2518–2528.

Grün, D., Kester, L., and van Oudenaarden, A. (2014a). Validation of noise models for single-cell transcriptomics. Nat. Methods *advance online publication*.

Grün, D., Kester, L., and van Oudenaarden, A. (2014b). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Grun, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature *525*, 251–255.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell *18*, 675–685.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell *146*, 633–644.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-seq: single-cell rna-seq by multiplexed linear amplification. Cell Rep. *2*, 666–673.

Hellwig, B., Hengstler, J.G., Schmidt, M., Gehrmann, M.C., Schormann, W., and Rahnenführer, J. (2010). Comparison of scores for bimodality of gene expression distributions and genome-wide evaluation of the prognostic relevance of high-scoring genes. BMC Bioinformatics *11*, 276.

Hogquist, K.A., Baldwin, T.A., and Jameson, S.C. (2005). Central tolerance: learning self-control in the thymus. Nat. Rev. Immunol. *5*, 772–782.

Hosmer, Jr., D.W., Lemeshow, S., and Sturdivant, R.X. (2013). Applied Logistic Regression (John Wiley & Sons, Inc.).

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biom. J. *50*, 346–363.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Huang, J., Lu, M.M., Cheng, L., Yuan, L.-J., Zhu, X., Stout, A.L., Chen, M., Li, J., and Parmacek, M.S. (2009c). Myocardin is required for cardiomyocyte survival and maintenance of heart function. Proc. Natl. Acad. Sci. *106*, 18734–18739.

Islam, S., Kjällquist, U., Moliner, A., Zajac, P., Fan, J.-B., Lönnerberg, P., and Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. *21*, 1160–1167.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods *11*, 163–166.

Jaeger, S., Fernandez, B., and Ferrier, P. (2013). Epigenetic aspects of lymphocyte antigen receptor gene rearrangement or “when stochasticity completes randomness”. Immunology *139*, 141–150.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. Science *343*, 776–779.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res. *21*, 1543–1551.

Kajimura, S., Seale, P., and Spiegelman, B.M. (2010). Transcriptional Control of Brown Fat Development. Cell Metab. *11*, 257–262.

Kamme, F., Salunga, R., Yu, J., Tran, D.-T., Zhu, J., Luo, L., Bittner, A., Guo, H.-Q., Miller, N., Wan, J., et al. (2003). Single-Cell Microarray Analysis in Hippocampus CA1: Demonstration and Validation of Cellular Heterogeneity. J. Neurosci. *23*, 3607–3615.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kaufmann, B.B., and van Oudenaarden, A. (2007). Stochastic gene expression: from single molecules to the proteome. Curr. Opin. Genet. Dev. *17*, 107–112.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. Nat. Methods *11*, 740–742.

Kim, J., and Eberwine, J. (2010). RNA: state memory and mediator of cellular phenotype. Trends Cell Biol. *20*, 311–318.

Kim, J.K., and Marioni, J.C. (2013). Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. Genome Biol. *14*, R7.

Kim, K.-T., Lee, H.W., Lee, H.-O., Kim, S.C., Seo, Y.J., Chung, W., Eum, H.H., Nam, D.-H., Kim, J., Joo, K.M., et al. (2015). Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells. Genome Biol. *16*.

Kim, T.K., Sul, J.-Y., Peternko, N.B., Lee, J.H., Lee, M., Patel, V.V., Kim, J., and Eberwine, J.H. (2011). Transcriptome transfer provides a model for understanding the phenotype of cardiomyocytes. Proc. Natl. Acad. Sci. U. S. A. *108*, 11918–11923.

Kumar, R.M., Cahan, P., Shalek, A.K., Satija, R., Jay DaleyKeyser, A., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J.J., et al. (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature *516*, 56–61.

Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefanik, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J., et al. (2013). Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell *153*, 976–987.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*.

Lee, J., Lee, J., Farquhar, K.S., Yun, J., Frankenberger, C.A., Bevilacqua, E., Yeung, K., Kim, E.-J., Balázsi, G., and Rosner, M.R. (2014a). Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. Proc. Natl. Acad. Sci. *111*, E364–E373.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., et al. (2014b). Highly Multiplexed Subcellular RNA Sequencing in Situ. Science *343*, 1360–1363.

Lee, M.-C.W., Lopez-Diaz, F.J., Khan, S.Y., Tariq, M.A., Dayn, Y., Vaske, C.J., Radenbaugh, A.J., Kim, H.J., Emerson, B.M., and Pourmand, N. (2014c). Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by RNA sequencing. Proc. Natl. Acad. Sci. *111*, E4726–E4735.

Levsky, J.M., and Singer, R.H. (2003). Gene expression and the myth of the average cell. Trends Cell Biol. *13*.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*.

Ligges, U., Maechler, M., and Schnackenberg, S. (2014). scatterplot3d: 3D Scatter Plot.

Liu, C., Maejima, T., Wyler, S.C., Casadesus, G., Herlitze, S., and Deneris, E.S. (2010). Pet-1 is required across different stages of life to regulate serotonergic function. Nat. Neurosci. *13*, 1190–1198.

Lorenz, R., Bernhart, S.H., Siederdissen, C.H. zu, Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol. Biol. *6*, 26.

Losick, R., and Desplan, C. (2008). Stochasticity and cell fate. Science *320*, 65–68.

Lovatt, D., Ruble, B.K., Lee, J., Dueck, H., Kim, T.K., Fisher, S., Francis, C., Spaethling, J.M., Wolf, J.A., Grady, M.S., et al. (2014). Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. Nat. Methods *11*, 190–196.

Majewski, J., and Pastinen, T. (2011). The study of eQTL variations by RNA-seq: from SNPs to phenotypes. Trends Genet. TIG *27*, 72–79.

Marder, E. (2011). Variability, compensation, and modulation in neurons and circuits. Proc. Natl. Acad. Sci. U. S. A. *108*, 15542–15548.

Marder, E., and Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. Nat. Rev. Neurosci. *7*, 563–574.

Martins, B.M., and Locke, J.C. (2015). Microbial individuality: how single-cell heterogeneity enables population level strategies. Cell Regul. *24*, 104–112.

McDavid, A., Dennis, L., Danaher, P., Finak, G., Krouse, M., Wang, A., Webster, P., Beechem, J., and Gottardo, R. (2014). Modeling Bi-modality Improves Characterization of Cell Cycle on Gene Expression in Single Cells. PLoS Comput Biol *10*, e1003696.

Miller, B.H., McDearmon, E.L., Panda, S., Hayes, K.R., Zhang, J., Andrews, J.L., Antoch, M.P., Walker, J.R., Esser, K.A., Hogenesch, J.B., et al. (2007). Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. Proc. Natl. Acad. Sci. *104*, 3342–3347.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res. *38*, D750.

Mineta, K., Matsumoto, T., Osada, N., and Araki, H. (2015). Population genetics of non-genetic traits: Evolutionary roles of stochasticity in gene expression. Gene *562*, 16–21.

Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O., and Lai, E.C. (2013). Widespread and extensive lengthening of 3' UTRs in the mammalian brain. Genome Res. *23*, 812–825.

Miyashiro, K., Dichter, M., and Eberwine, J. (1994). On the nature and differential distribution of mRNAs in hippocampal neurites: implications for neuronal functioning. Proc. Natl. Acad. Sci. *91*, 10800–10804.

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A.J., Tanaka, Y., Wilkinson, A.C., Buettner, F., Macaulay, I.C., Jawaid, W., Diamanti, E., et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nat Biotech *33*, 269–276.

Morgan, M., Pages, H., and Obenchain, V. Rsamtools: Binary alignment (BAM), variant call (BCF), or tabix file import.

Morris, J., Singh, J.M., and Eberwine, J.H. (2011). Transcriptome analysis of single cells. J. Vis. Exp.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Munsky, B., Neuert, G., and Oudenaarden, A. van (2012). Using gene expression noise to understand gene regulation. Science *336*, 183–187.

Muotri, A.R., Chu, V.T., Marchetto, M.C.N., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903–910.

Muotri, A.R., Marchetto, M.C.N., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443–446.

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., Oles, A.K., Arauzo-Bravo, M.J., Saitou, M., Hadjantonakis, A.-K., et al. (2014). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. Nat. Cell Biol. *16*.

O'Leary, T., Williams, A.H., Caplan, J.S., and Marder, E. (2013). Correlations in ion channel expression emerge from homeostatic tuning rules. Proc. Natl. Acad. Sci. *110*, E2645–E2654.

O'Rourke, N.A., Weiler, N.C., Micheva, K.D., and Smith, S.J. (2012). Deep molecular diversity of mammalian synapses: why it matters and how to measure it. Nat Rev Neurosci *13*, 365–379.

Overton, K.W., Spencer, S.L., Noderer, W.L., Meyer, T., and Wang, C.L. (2014). Basal p21 controls population heterogeneity in cycling and quiescent cell cycle states. Proc. Natl. Acad. Sci. *111*, E4386–E4393.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. *12*.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol. Cell *58*, 339–352.

Park, J., Brureau, A., Kernan, K., Starks, A., Gulati, S., Ogunnaike, B., Schwaber, J., and Vadigepalli, R. (2014). Inputs drive cell phenotype variability. Genome Res. *24*, 930–941.

Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

Patil, S., Fribourg, M., Ge, Y., Batish, M., Tyagi, S., Hayot, F., and Sealfon, S.C. (2015). Single-cell analysis shows that paracrine signaling by first responder cells shapes the interferon-beta response to viral infection. Sci. Signal. *8*.

Paulsson, J. (2005). Models of stochastic gene expression. Phys. Life Rev. *2*, 157–175.

Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods *10*, 1096–1098.

Pipes, G.C.T., Creemers, E.E., and Olson, E.N. (2006). The myocardin family of transcriptional coactivators: versatile regulators of cell growth, migration, and myogenesis. Genes Dev. *20*, 1545–1556.

Piras, V., Tomita, M., and Selvarajoo, K. (2014). Transcriptome-wide variability in single embryonic development cells. Sci. Rep. *4*.

Ponting, C.P., and Belgard, T.G. (2010). Transcribed dark matter: meaning or myth? Hum. Mol. Genet. *19*, R162–R168.

Poulin, J.-F., Zou, J., Drouin-Ouellet, J., Kim, K.-Y.A., Cicchetti, F., and Awatramani, R.B. (2014). Defining Midbrain Dopaminergic Neuron Diversity by Single-Cell Gene Expression Profiling. Cell Rep. *9*, 930–943.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biol *4*, e309.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Raj, A., Rifkin, S.A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. Nature *463*, 913–918.

Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol *5*, e1000598.

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. *30*, 777–782.

Rand, U., Rinas, M., Schwerk, J., Nohren, G., Linnes, M., Kroger, A., Flossdorf, M., Kaly-Kullai, K., Hauser, H., Hofer, T., et al. (2012). Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. Mol. Syst. Biol. *8*.

Raser, J.M., and O'Shea, E.K. (2005). Noise in gene expression: origins, consequences, and control. Science *309*, 2010–2013.

R Development Core Team (2010). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

Rifkin, S.A., Atteson, K., and Kim, J. (2000). Constraint structure analysis of gene expression. Funct. Integr. Genomics *1*, 174–185.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., et al. (2008). Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell *132*, 487–498.

Sansom, S.N., Shikama-Dorn, N., Zhanybekova, S., Nusspaumer, G., Macaulay, I.C., Deadman, M.E., Heger, A., Ponting, C.P., and Holländer, G.A. (2014). Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. Genome Res. *24*, 1918–1931.

Sasagawa, Y., Nikaido, I., Hayashi, T., Danno, H., Uno, K.D., Imai, T., and Ueda, H.R. (2013). Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol. *14*, R31.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*.

Schulz, D.J., Goaillard, J.-M., and Marder, E.E. (2007). Quantitative expression profiling of identified neurons reveals cell-specific constraints on highly variable levels of gene expression. Proc. Natl. Acad. Sci. *104*, 13187–13191.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Scott, M.M., Wylie, C.J., Lerch, J.K., Murphy, R., Lobur, K., Herlitze, S., Jiang, W., Conlon, R.A., Strowbridge, B.W., and Deneris, E.S. (2005). A genetic approach to access serotonin neurons for in vivo and in vitro studies. Proc. Natl. Acad. Sci. U. S. A. *102*, 16472–16477.

SEQC/MAQC-III Consortium (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotech *32*, 903–914.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature *510*, 363–369.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. *14*, 618–630.

Sharman, J.L., Benson, H.E., Pawson, A.J., Lukito, V., Mpamhanga, C.P., Bombail, V., Davenport, A.P., Peters, J.A., Spedding, M., Harmar, A.J., et al. (2012). IUPHAR-DB: updated database content and new features. Nucleic Acids Res. *41*, D1083–D1088.

Singh, D.K., Ku, C.-J., Wichaidit, C., Steininger, R.J. 3rd, Wu, L.F., and Altschuler, S.J. (2010). Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. Mol. Syst. Biol. *6*.

Smith, C.L., and Eppig, J.T. (2009). The mammalian phenotype ontology: enabling robust annotation and comparative analysis. Wiley Interdiscip. Rev. Syst. Biol. Med. *1*, 390–399.

Snijder, B., and Pelkmans, L. (2011). Origins of regulated cell-to-cell variability. Nat Rev Mol Cell Biol *12*, 119–125.

Spaethling, J.M., Sanchez-Alavez, M., Lee, J., Xia, F.C., Dueck, H., Wang, W., Fisher, S.A., Sul, J.-Y., Seale, P., Kim, J., et al. (2015). Single-cell transcriptomics and functional target validation of brown adipocytes show their complex roles in metabolic homeostasis. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.

Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., and Sorger, P.K. (2009). Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature *459*, 428–432.

Sul, J.-Y., Wu, C. -w. K., Zeng, F., Jochems, J., Lee, M.T., Kim, T.K., Peritz, T., Buckley, P., Cappelleri, D.J., Maronski, M., et al. (2009). Transcriptome transfer produces a predictable cellular phenotype. Proc. Natl. Acad. Sci. *106*, 7624–7629.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

Tang, F., Lao, K., and Surani, M.A. (2011). Development and applications of single-cell transcriptome analysis. Nat. Methods *8*, S6–S11.

Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. Nature *466*, 267–271.

Tibshirani, R., and Hastie, T. (2007). Outlier sums for differential gene expression analysis. Biostatistics *8*, 2–8.

Tietjen, I., Rihel, J.M., Cao, Y., Koentges, G., Zakhary, L., and Dulac, C. (2003). Single-Cell Transcriptional Analysis of Neuronal Progenitors. Neuron *38*, 161–175.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Usoskin, D., Furlan, A., Islam, S., Abdo, H., Lonnerberg, P., Lou, D., Hjerling-Leffler, J., Haeggstrom, J., Kharchenko, O., Kharchenko, P.V., et al. (2015). Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. Nat Neurosci *18*, 145–153.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Natl. Acad. Sci. *87*, 1663–1667.

Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. Mol. Cell *47*, 827–838.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. *10*.

Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sanchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., et al. (2015). Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell *16*, 712–724.

Woods, H.A. (2014). Mosaic physiology from developmental noise: within-organism physiological diversity as an alternative to phenotypic plasticity and phenotypic flexibility. J. Exp. Biol. *217*.

Xue, Q., Lu, Y., Eisele, M.R., Sulistijo, E.S., Khan, N., Fan, R., and Miller-Jensen, K. (2015). Analysis of single-cell cytokine secretion reveals a role for paracrine signaling in coordinating macrophage responses to TLR4 stimulation. Sci. Signal. *8*.

Zheng, W., Chung, L.M., and Zhao, H. (2011). Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics *12*, 290.

Zhou, D., Wang, Y., and Wu, B. (2014). A multi-phenotypic cancer model with cell plasticity. J. Theor. Biol. *357*.