ANNUAL REVIEWS

Annual Review of Genomics and Human Genetics

Using Single-Cell and Spatial Transcriptomes to Understand Stem Cell Lineage Specification During Early Embryo Development

Guangdun Peng,[1,2,3] Guizhong Cui,[2] Jincan Ke,[1] and Naihe Jing[1,2,3,4]

[1]CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China; email: peng_guangdun@gibh.ac.cn

[2]Center for Cell Lineage and Atlas, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou 510005, China

[3]Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

[4]State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, University of the Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China; email: njing@sibcb.ac.cn

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

ANNUAL REVIEWS CONNECT
www.annualreviews.org
• Download figures
• Navigate cited references
• Keyword search
• Explore related articles
• Share via email or social media

Annu. Rev. Genom. Hum. Genet. 2020. 21:163–81

First published as a Review in Advance on April 27, 2020

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

https://doi.org/10.1146/annurev-genom-120219-083220

Copyright © 2020 by Annual Reviews.
All rights reserved

Keywords

single-cell genomics, spatial transcriptome, cell lineage, embryo development, trajectory inference

Abstract

Embryonic development and stem cell differentiation provide a paradigm to understand the molecular regulation of coordinated cell fate determination and the architecture of tissue patterning. Emerging technologies such as single-cell RNA sequencing and spatial transcriptomics are opening new avenues to dissect cell organization, the divergence of morphological and molecular properties, and lineage allocation. Rapid advances in experimental and computational tools have enabled researchers to make many discoveries and revisit old hypotheses. In this review, we describe the use of single-cell RNA sequencing in studies of molecular trajectories and gene regulation networks for stem cell lineages, while highlighting the integrated

experimental and computational analysis of single-cell and spatial transcriptomes in the molecular annotation of tissue lineages and development during postimplantation gastrulation.

INTRODUCTION

Stem cells and stem cell–based regenerative medicine represent the frontier of next-generation health care and hold the promise of curing diseases by repairing, replacing, engineering, or restoring damaged tissues or organs. Given that there are still obstacles in producing functional and mature cell types efficiently and precisely, a better understanding of normal developmental and differentiation processes will lay the foundation for generating authentic cell types for translational applications and drug screening.

Conventionally, the study of stem cell fate determination of embryos has relied on tracking cell movement or genetic perturbation of a few molecules at multiple developmental time points (46). However, there has been no complete and detailed map of all cell types in the human body and the developmental connectivity among them, as there is for *Caenorhabditis elegans*, which has a lineage tree for precisely 671 cells (11). Moreover, it is difficult to probe the biological events due to the sparsity of the pluripotent cells and the dynamic nature of development. The allocation of progenitor cells appears to be profoundly influenced by the positions of these cells in the embryo. Cells are endowed with a specific lineage potency in the context of complicated cell–cell interactions and morphogen gradients.

Innovative technological breakthroughs bring innovative discoveries and new findings. Instead of relying on a large number of cells (in which signals are masked by bulk populations), single-cell RNA sequencing (scRNA-seq), with its ability to assay properties of individual cells and align cells of continuously changing status, has helped decipher lineage commitment for stem cell differentiation and embryo development (32, 91) and has proved to be a key technical platform for biomedical research (reviewed in 25, 44, 60, 82, 122).

CELL FATE DETERMINATION AND LINEAGE SPECIFICATION OF THE GASTRULATION EMBRYO

The paradigm and blueprint of lineage segregation in early mammalian embryos are established during gastrulation, in which progenitors of various cell fates are regionalized and patterned in different embryo locations. Studies of the developmental anatomy of vertebrate embryos point to the existence of a phylotypic stage of development when the embryos of different species display a basic body plan (35). Fate mapping and lineage analysis of the gastrula-stage embryos of zebrafish, *Xenopus*, chicks, and mice have revealed a similar pattern of cell movement and regionalization of cell fates among these species (83, 93, 110), reinforcing the concept that the primary outcome of early embryogenesis is the establishment of a basic body plan on which the whole embryo is built for embryonic tissue patterning. Gastrulation is therefore a choreographed sequence of cell fate decisions, cell proliferation, and cell movement and a milestone during early embryo development, through which the multipotent stem cells in the epiblast are spatially organized and orchestrated to give rise to diverse cell types (7, 94, 109).

Importantly, early embryo development in model organisms exhibits common molecular features, including the signaling landscape, transcriptional regulation, and germ layer signatures, as also evidenced in recent in vitro embryo cultures from mice, monkeys, and humans (59, 73, 124, 125). The remarkable similarity in the stem cell behavior during periods of early embryonic development points to the potential existence of a coherent principle and orchestration of

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

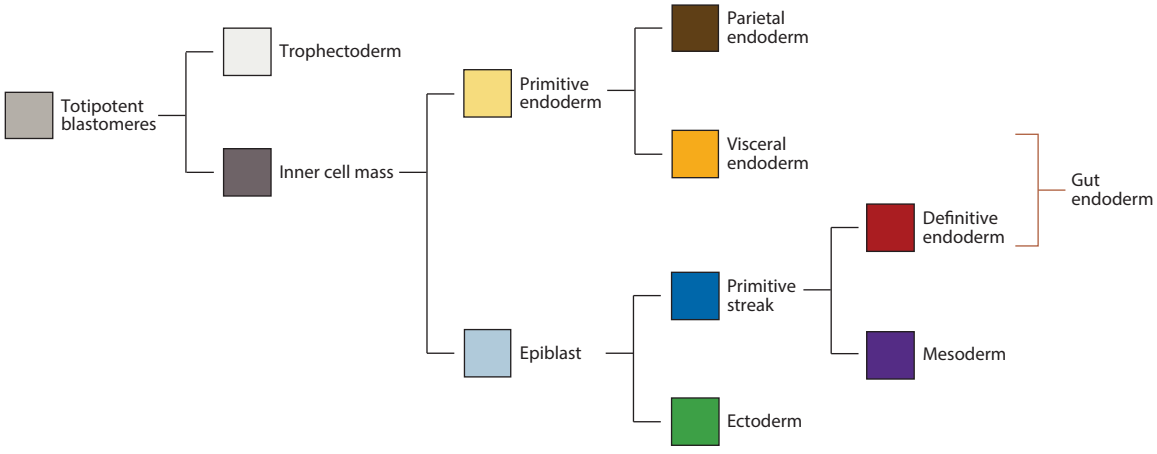

Figure 1

Cell lineage specification in early embryo development. After fertilization, cells divide as totipotent blastomeres. At approximately embryonic day 2.5 (E2.5), the first cell fate decision in the mouse embryo takes place when two populations of cells are segregated: the trophectoderm and the inner cell mass. In the second fate decision at E3.5, cells of the inner cell mass differentiate into the primitive endoderm and epiblast. Around implantation (E4.5), the primitive endoderm differentiates into the parietal endoderm and the visceral endoderm. In the classical view, the trophectoderm and primitive endoderm contribute to extraembryonic tissue; however, studies have found that some of the derivatives of the visceral endoderm contribute to the gut endoderm. At E6.5, gastrulation occurs at the primitive streak, and posterior epiblast cells, through an epithelial–mesenchymal transition, segregate into the mesoderm and definitive endoderm, while the rest of the epiblast develops into the ectoderm.

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

transcriptional regulation underpinning the lineage specification, which governs the establishment of the stem cell hierarchy and ensures the stepwise control of development and differentiation, through which the encoding and coordinating morphogenetic outcomes are achieved (37, 38, 53). Remarkably, the scale and complexity of the mouse gastrulation embryo, at a size of hundreds of micrometers and 500–15,000 cells (43), also provide an attainable experimental system for dissecting the single-cell and spatial organization of tissue progenitors (**Figure 1**).

UNDERSTANDING A CELL'S BEHAVIOR

Since the invention of the microscope by the Dutch scientist Antonie van Leeuwenhoek in 1660, scientists have gained a deep understanding of the microcosm of a cell. It was not until 1855 that the German scientist and physician Rudolf Virchow popularized the cell theory—that all living organisms are made of cells, that cells are the basic unit of life for all living things, and that all cells arise from preexisting cells—and people began to explain diseases in terms of cell abnormalities (63). However, due to various technical obstacles, the comprehensive profiling of the transcriptional activities of individual cells was not effectively utilized until the emergence of high-throughput sequencing in 2005 and the first report of scRNA-seq in 2009 (111). Two years later, Navin et al. (71) of Cold Spring Harbor Laboratory completed the world's first single-cell DNA sequencing. This pioneering research directly initiated the new era of single-cell genomics.

With its high resolution and high sensitivity, single-cell genomics allows scientists to examine subtle differences among individual cells at a molecular level. Rather than surveying many mixed cells, which masks the individual types, exploration of single cells provides an important way to probe deep differences in many dimensions, turning biological processes into mathematic probabilities or statistics, which has transformed many disciplines of biomedical research (80, 91).

Single-cell sequencing has undergone several generations of technological developments as well as major breakthroughs in scalability and affordability. It has also expanded from the transcriptome to the genome, the epigenome, and even the proteome. Here, we focus on single-cell transcriptomic studies, as the transcriptome represents the phenotypic output of a cell and has been the most studied area at present. In particular, the introduction of popular commercial platforms such as those of Fluidigm and 10x Genomics has made it significantly easier to carry out scRNA-seq. Single-cell analysis has become a routine tool in laboratories for studying tumor biology, neurology, stem cells, and developmental biology (49).

DIMENSIONALITY REDUCTION ANALYSIS FOR SINGLE-CELL RNA DATA

Compared with bulk RNA-seq, scRNA-seq represents much higher dimensions in both cell numbers and the complexity of gene expression, such as sparsity or dropout. Dimensionality reduction and visualization are therefore the primary tasks for scRNA-seq data analysis. A quick look into the data using principal component analysis (PCA) to identify cell clusters with similar gene expression profiles could reveal the basic features of data structure and allow a rough evaluation of sample coverage or batch interference. For instance, outlier cell populations may arise from low-quality sequencing or bad sample preparation. As a linear reduction method, the major principal components may explain the key biological differences between cell clusters, since they preserve the distance structure within the data. Generally, the top two or three principal component spaces are the most informative. A three-dimensional PCA or display on other principal component spaces could also be meaningful in revealing biological significance. The top PCA loading genes or significantly enriched genes along the principal component axis may account for the variances for discriminating cell distribution, which could be further used for iterative PCA clustering.

By comparison, *t*-distributed stochastic neighbor embedding (t-SNE) is a nonlinear dimensionality reduction method that tends to preserve local distances over global distance (117). t-SNE is an efficient way to identify the variations and is popular in processing single-cell data, especially when the heterogeneity of single cells is not readily detectable. However, it is not currently possible to explain variances in a t-SNE map. As a combination of features from t-SNE and PCA, uniform manifold approximation and projection (UMAP) provides specific, concrete visualization of the data in addition to preserving the global structures (9).

Dimensionality reduction continues to be a fast-developing research field (101, 102), and an emerging trend is the involvement of machine learning or deep learning approaches, which are implemented to model and estimate the data distribution. With their ability to learn and summarize data features through encoders and, importantly, to impute missing data values and predict new biological patterns via decoders, new algorithms have been developed to mitigate the challenges of single-cell data analysis. For example, Lin et al. (55) used a neural network for dimensionality reduction and to infer cell types in scRNA-seq data by taking advantage of its ability to learn different combinations of gene expression levels. The autoencoder-based AutoImpute was developed to compensate for the missing values of scRNA-seq (105). Convolutional neural networks could effectively extract multiple-dimension spatial features. Recently, a residual fully connected neural network combined with hyperparameter search and efficient network architecture search was devised for analyzing data with a structured hierarchical network, which could facilitate end-to-end automated machine learning for scRNA-seq data (56). Similarly, the Sparse Autoencoder for Unsupervised Clustering, Imputation, and Embedding (SAUCIE) streamlines the processing of multisample data simultaneously under a single unified representation learned by a deep autoencoder, ensuring satisfactory performance in batch removal, imputation, clustering, and dimensionality reduction (5).

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

TRAJECTORY ANALYSIS FOR SINGLE-CELL RNA SEQUENCING DATA

In addition to clustering single cells, identifying cell type heterogeneity, and revealing the differential gene expression changes of major cell types, it is important to decipher the progressive status of single cells to reveal cellular decision-making. Conventionally, time-series data points or heritable marks followed by tracking are required to construct the cell lineage. A significant challenge is to obtain sufficient samples to cover enough changes or ensure labeling marks of high specificity. Cells within a population can exhibit temporal heterogeneity. They may, for example, differ primarily with regard to the stage (e.g., of a developmental process) at which they are sampled. With transcriptomic data from a large number of single cells representing a continuous spectrum of cell states [e.g., by droplet sequencing (Drop-seq), 10x Genomics, or split pool ligation-based transcriptome sequencing (SPLiT-seq)], pseudotemporal ordering algorithms have been developed to sort cells along the developmental trajectory and reveal the lineage relationship that is encoded in the distance of gene expression. The beauty of scRNA-seq trajectory inference is that lineage information is independent from the genetic relationships among cells and can be predicted even when cells are sampled only at a fixed time point (42). The history of a cell is therefore obtained by decoding the many time events in one moment, similar to William Blake's line "To see a World in a Grain of Sand." This is particularly relevant for scenarios in which the past status of a sample cannot be obtained, as is the case in many pathological or developmental processes. Needless to say, this technique allows the identification of cell types along the pseudotime representing early, intermediate, and end points as well as the characterization of turning factors that accompany the trajectory (27) (**Figure 2**).

Several computational tools based on this theoretic context have been developed to model the differentiation process and single-cell behaviors in scRNA-seq data (86). They all consist of two steps for dimension reduction and ordering. For example, Monocle places cells in a low-dimensional space and constructs a minimum spanning tree to produce the developmental trajectory (114). SCORPIUS infers trajectories in a purely data-driven way, without prior information about the dynamic process (12). Slingshot takes advantage of the combined techniques for identifying multiple branching lineages (100). Partition-based graph abstraction (PAGA) preserves both continuous cell transitions and discrete cell types by generating graph-like maps of cells (121). Harmony combined with the Palantir algorithm takes defined starter cells based on prior knowledge to build a unidirectional trajectory (75).

Except for the transcription information, additional data layers for predictive trajectory models should also be superimposed. Integrating different levels of information, such as cell labeling, experimental time, the delineation of mRNA from precursor mRNA (48), and population size measurements, can improve the accuracy of inference trajectories.

It is worth noting that, although there are many approaches to trajectory inference, there is no single best approach for all types of data sets (115). In addition, the functions of models for reconstructing cell trajectories are complementary. Different methods should be tested when dealing with different data sets and considering the composite complexity of biological trajectories.

INFERRING CELL LINEAGES USING SYSTEMATIC SINGLE-CELL RNA SEQUENCING IN EMBRYO DEVELOPMENT

Single-cell data have highlighted the continuous nature of cellular phenotypes and provided a probabilistic framework to elucidate cell fate decision-making and delineate the transition paths and intermediate states among all cell types based on gene similarity. Several recent studies have taken advantage of systematic single-cell transcriptomes to build a lineage tree. For example, using Drop-seq, Farrell et al. (23) sequenced a large number of single cells with high temporal

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

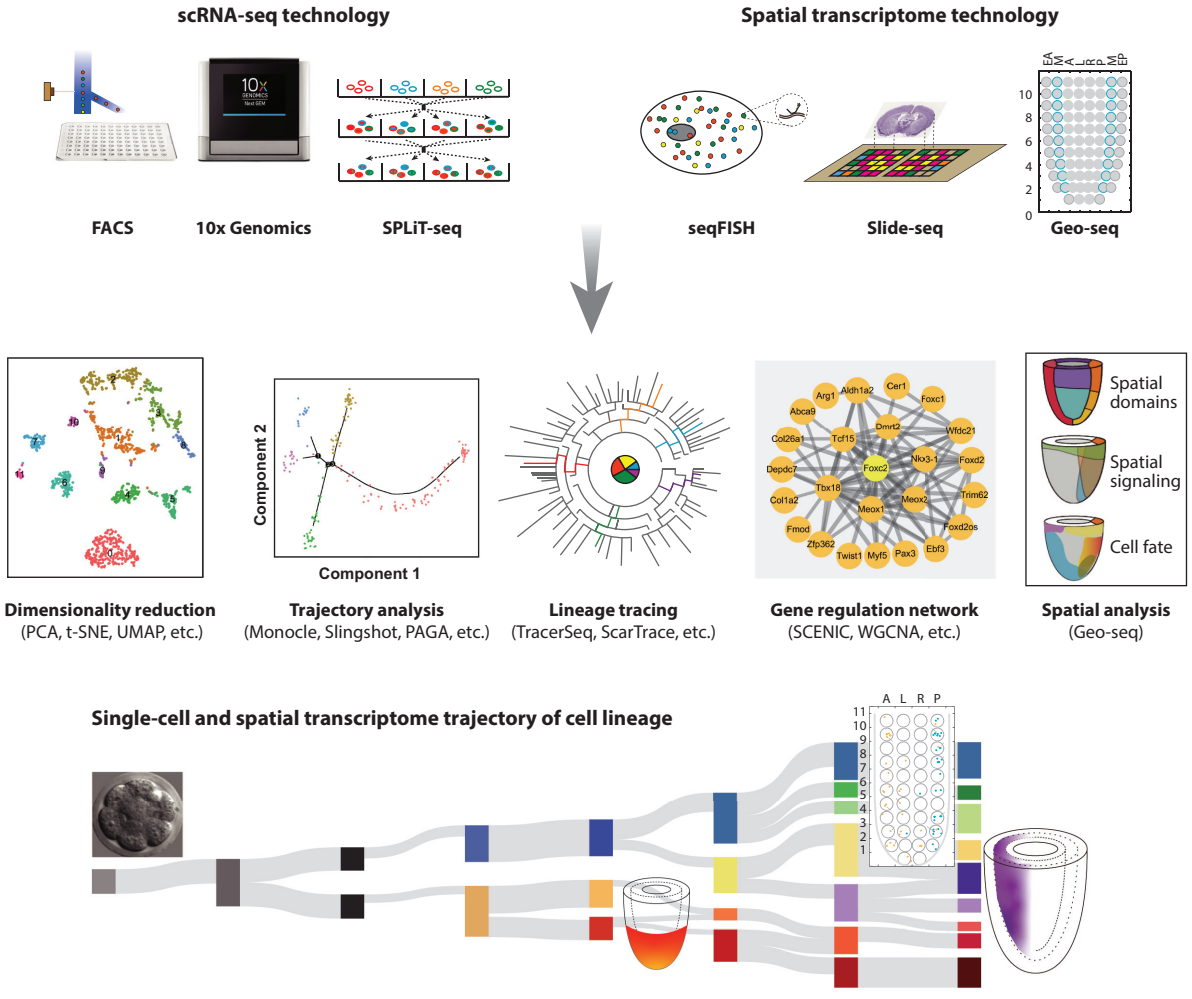

Figure 2

Single-cell and spatial transcriptome approaches to mapping cell lineages. scRNA-seq and spatial transcriptome technologies are used to perform dimensionality reduction analysis, trajectory analysis, and regulatory network analysis in order to infer a common mechanism of lineage specification. Mouth pipetting and FACS have been used for small populations of rare cells. The Drop-seq, 10x Genomics, and SPLiT-seq methods were developed for unsupervised scRNA-seq and provide improved throughput. Direct and indirect spatial transcriptome technologies (seqFISH, Slide-seq, and Geo-seq) provide cell spatial information, and single-cell clonal-tracking methods (TracerSeq and ScarTrace) record the cell history. These approaches could enable researchers to identify cell lineages at single-cell and spatial levels. Abbreviations: Drop-seq, droplet sequencing; FACS, flow-activated cell sorting; Geo-seq, geographical position sequencing; PAGA, partition-based graph abstraction; PCA, principal component analysis; SCENIC, single-cell regulatory network inference and clustering; scRNA-seq, single-cell RNA sequencing; seqFISH, sequential fluorescence in situ hybridization; Slide-seq, slide sequencing; SPLiT-seq, split pool ligation-based transcriptome sequencing; t-SNE, *t*-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection; WGCNA, weighted gene coexpression network analysis.

resolution to uncover transcriptional trajectories during zebrafish embryogenesis. They developed a simulating diffusion-based reconstruction method, URD (named after one of the mythological Norse figures who decide human fates), which utilizes discrete random walks toward the root and graph searches on biased transition probabilities to estimate the continuous process of diffusion. Because URD requires user-defined starting points and end points, this computational

method involves more biological developmental information and is a powerful way to visualize the branching tree of developmental trajectories. Nowotschin et al. (75) also took advantage of the graph-based approaches to model different endoderm cells within a mouse embryo and developed a Harmony algorithm (45) to bridge the developmental time points. Harmony uses mutual nearest neighbors to construct an augmented *k* nearest-neighbor graph that connects time points. The Palantir algorithm infers cell fate potential, which takes as input a user-defined early root and infers pseudotime and branch probabilities; it models cell fate choices as continuous probabilistic processes rather than treating lineage decisions as bifurcations. Integrating Harmony with Palantir enabled Nowotschin et al. (75) to construct a spatiotemporal trajectory map of the developing endoderm. Their study characterized the bifurcation and convergence events for gut endoderm formation from both embryonic and extraembryonic origins.

However, the path for stem cell fate determination could be much more complicated. The developmental trajectory is not restricted to linear or bifurcating structures, and more than one branch or multiple graph topologies could be intermingled in the development and differentiating process (86). Moreover, discontinuous cell states or abrupt transcriptomic changes can occur, such as stem cell asymmetrical division. The current trajectory inference methods are limited in reconstructing lineages with multiple disconnected topologies or cycles or complex tree structures (86). They also take the input of the transcriptome as the cell phenotype, but factors other than the transcriptome, such as metabolic, epigenetic, and splicing regulation, are also vital for lineage differentiation (41, 54, 112, 123). Therefore, enhanced methods suitable for complex trajectories and larger numbers of cells and features are needed to improve lineage reconstructions and make them more precise.

COMBINING SINGLE-CELL RNA SEQUENCING WITH LINEAGE TRACING

One of the central questions in developmental biology and stem cell biology is how a fertilized egg develops into the various cell types that make up the fully formed organism, i.e., the cell lineage. Using a static continuum of cell states to interpret a dynamic trajectory remains challenging even with trajectory analysis on single cells. With the incorporation of cell history from single-cell clonal tracking (in organoids or model organisms or endogenously) and cell type information gleaned from single-cell transcriptomes, a higher-fidelity cell lineage could be derived along with a cell fate map.

Inspired by this idea, the genomic DNA barcode was originally introduced with viral transfection to the founder cells of hematopoietic stem cells, followed by scRNA-seq (57). Wagner et al. (120) developed a transposon-based barcoding approach, TracerSeq, for reconstructing the single-cell lineage history. This technique uses the Tol2 transposon to randomly incorporate a uniquely barcoded GFP reporter into the zebrafish genome for clone marking and tracking. Besides TracerSeq, Cas9 nucleases can be used to target a specific locus and generate random insertions or deletions (indels) through DNA repair mechanisms. These indels are inheritable and can therefore serve as genetic tracers for lineage tracing. For example, some researchers have used CRISPR/Cas9-based technologies to record cell lineages on one-cell-stage embryos and combined this approach with scRNA-seq to generate fate maps in zebrafish (96). Alemany et al. (4) recently devised ScarTrace on a transgenic zebrafish line with multiple copies of histone GFP. Cas9 and guide RNA–targeting GFP were injected into the zygote cell to introduce multiple indels. Similarly, to increase the complexity of CRISPR barcodes, a mouse model carrying a self-targeting form of Cas9 has been created to study the lineage contributions of major organs (36). To further increase the barcode complexity, Chan et al. (14) developed a high-information,

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

multichannel molecular recorder as an evolving lineage tracer to map cell fate transitions from fertilization through gastrulation. The implementation of a genomic DNA barcode allows cell labeling of unlimited complexity, thus greatly expanding the throughput of lineages that can be traced simultaneously. Theoretically, each of the individual cells of the whole body can be barcoded, providing unprecedented coverage for lineage tracing. Along with the single-cell transcriptomic profiling, the lineage and fate relations of cells are mingled together, providing not only an informative lineage history but also a powerful validation of cell types and transitions. Therefore, this large-scale cellular genetic barcoding combined with scRNA-seq showcases the power of unbiased organism-wide lineage tracing and cell fate mapping (8).

The DNA methylation and hydroxymethylation status in single cells is also inheritable and allows precise cell lineage reconstruction. Importantly, epigenetic modifications provide a unique lens on cellular state, which can be more robust across time than the dynamic transcriptome, and often convey more detailed mechanistic information as well (68).

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

GENE REGULATION NETWORKS REVEALED BY SINGLE-CELL RNA SEQUENCING

Coupling information from different expression modalities revealed by scRNA-seq and the developmental trajectories could enable the delineation of a cell hierarchy or intermediate cell states and unveil transcriptional regulatory mechanisms. To this end, a variety of computational and experimental methods for inferring gene regulation networks (GRNs) with single-cell data have been developed. For instance, to reveal the significance of GRNs from single-cell data, single-cell regulatory network inference and clustering (SCENIC) links the coexpression modules with *cis*-regulatory sequences and defines the cell states via binarization of the single-cell data. The regulon activities are scored in each cell as replacements for gene expression profiles, and the GRN based on the scored regulon not only facilitates batch-insensitive data integration but also enhances the mechanistic interpretation of the transcriptomic data because of the inclusion of motif-associated regulatory information (3). This analysis has identified the cell type–specific regulatory network structure and critical transcription regulators for the maintenance of cell identity from a mouse single-cell atlas data set and aging *Drosophila* brain (19, 103). In comparison with transcriptome-level analysis, the GRN-based approaches can also provide consistent performance when applied to complex population data. For example, the spatial and temporal dynamics of GRNs for tissue architecture have been established for the postimplantation mouse embryo, unveiling the stepwise regulatory programs in cell lineage determination (79). Regulon activity can be further combined with traditional analytical approaches, such as weighted gene coexpression network analysis (50), dimensionality reduction, and the connection specificity index, to uncover potential functional modules and define a matrix of regulon pairs with similar profiles (24), where the edges or nodes between regulons in the matrix denote potential positive or negative interactions. Using tools from graph theory to study the global and local properties of networks, a computational analytical framework that computes gene-to-gene correlation metrics from large-scale single-cell data enables a comprehensive characterization of GRNs (33).

DIRECT SPATIAL TRANSCRIPTOME ANALYSIS

Current single-cell sequencing technology needs to first decompose the solid tissues into individual single cells, a process that inevitably results in the loss of spatial information. Importantly, the dissociation procedure when preparing a single-cell suspension often takes time and requires multiple steps, making the stability and fidelity of RNA transcripts problematic (116). In situ spatial

transcriptome analysis aims to survey the natural state of cells in native tissues, in order to identify the location-defined cell types and understand how the cells are communicating with their niche, and therefore includes many important biological values.

Spatial transcriptome analysis is rapidly evolving, and the current technologies have an obvious trade-off between resolution and throughput. Traditionally, the integration of gene expression and spatial information in biological studies has relied on in situ methods, such as in situ hybridization or immunohistological staining. Although these techniques can achieve cellular or subcellular resolution, they cannot be used to examine the molecular complexity of a cell because they require predefined markers based on prior knowledge, and the number of markers that they can simultaneously assay is limited (18). The endeavor to improve the detection throughput includes adding the temporal or spectral conditions during the hybridization. For example, Lubeck et al. (58) developed sequential fluorescence in situ hybridization (seqFISH), a single-molecule method that uses multiple rounds of hybridization and rehybridization after bleaching or stripping probes as a way to define transcripts through the combination of fluorescent color. SeqFISH can simultaneously detect the in situ hybridization signals of hundreds of genes, greatly improving the throughput. Recently, the same laboratory used extended pseudocolor readout probes to perform successive rounds of in situ hybridizations and was able to detect 10,000 genes (22). Similarly, the multiplexed error-robust fluorescence in situ hybridization (MERFISH) technique utilizes additional spectral information, a combination of targeting sequence and readout sequence, to achieve high levels of multiplexing in single cells. The detection accuracy can be further enhanced by an error-correction decoding algorithm (16, 67).

Unlike methods based on in situ hybridization, the array-based spatial transcriptome protocol developed by Ståhl et al. (97) enables examination of the locations of hundreds of gene targets in intact tissue slices. Immersing brain slices in an array with anchored spatial-barcoded primers further permeates them, so that the mRNA in the cells binds directly to the anchor primer. Barcoded cDNA is then cleaved, prepared into libraries, and sequenced on a standard Illumina platform. Each barcoded RNA is then mapped back to its spatial location within 1 of ~1,000 different molecular positions, enabling in situ transcriptome sequencing on brain sections at a resolution of 100 μm. Additional platforms, such as slide sequencing (Slide-seq) (85) and high-definition spatial transcriptomics (HDST) (118), have emerged to provide the positional context of genome-wide expression in intact tissue sections with nearly single-cell resolution and high cell density coverage.

However, multiplex fluorescent FISH is dependent on highly sensitive probes, a customized gene set, and a super-resolution microscope. In addition, for methodologies that can measure hundreds of transcripts simultaneously, molecular crowding due to high-abundance transcripts within cells and tight contact between cells would result in a fluorescence/signal overlap, posing significant challenges for microscopy and computational deconvolution (52). The array-based in situ sequencing method has a limited gene detection ability and insufficient flexibility, and the success rates so far are not ideal (62). Although the density and resolution have been greatly improved by Slide-seq and HDST, both methods are currently limited by their sparse data signals and complicated confounding noise.

Geographical position sequencing (Geo-seq) with the integration of laser capture microdissection and scRNA-seq is a laborious but high-resolution approach that allows researchers to study the transcriptomes in as few as five cells from defined geographical locations. It has the advantage of very high gene detection efficiency and direct location preservation. With the accompanying spatial information, a three-dimensional transcriptome atlas can be built up to display the transcriptome spatially and quantitatively (15, 78). The transcriptomic analysis of embryonic tissues in real time and space provides an authentic benchmark for dissecting the cell organization,

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

molecular architecture, and lineage allocation (79). The ability to discern spatial gene expression features in complex biological systems is crucial for deciphering the mechanisms of developmental biology and disease (10, 17, 39, 95).

INDIRECT SPATIAL TRANSCRIPTOME

In addition to using direct in situ single-cell spatial sequencing technology, one can use the spatial position omics data that have been obtained and extract the spatial position reference coordinate system to restore the spatial positions of single cells and indirectly increase the resolution of the spatial transcriptome to single cells. For example, Satija et al. (87) developed the Seurat method, which first divides the embryo or tissue into grids and then converts the expression of the in situ hybridization signal of the known specific landmarks in a specific grid into binary (either "has" or "none"), thereby obtaining spatial positioning reference coordinates. By aligning the expression of these spatial coordinate system genes from single cells to the tissue grids, one can calculate the possibility of positioning a single cell in a specific grid. However, the success rate of spatial localization is often adversely affected by the limited number of reference coordinate genes with positional information or the loss of part of the gene expression when these genes are expressed in a single cell. To overcome this problem, the Seurat method performs machine learning to determine the weights of all positional reference genes in each grid and then uses these converted values for spatial localization. This improved method has enhanced the accuracy of predicting single-cell locations (87). Similarly, Achim et al. (1, 2) developed a mapping approach based on a subset of positional reference genes that has in situ hybridization information and calculated specificity scores for each of these genes in each cell. Combined with each cell and reference position partition correlation coefficient, these scores gave the most likely positional relationship of individual cells. The DistMap pipeline also binarizes the single-cell data and spatially maps the single cells based on an existing in situ database (40).

When using scRNA-seq in cases where spatial information contributes significantly to the cellular states, the analogous approaches for ordering single cells in developmental pseudotime could potentially also uncover developmental pseudospace. Many pseudotime methodologies have been adjusted and tried. Using diffusion maps (28), Scialdone et al. (88) identified the diffusion-space direction of the anterior–posterior axis of the primitive streak, which they also interpreted as the pseudospatial coordinates representing the positions of cells along the primitive streak. Farrell et al. (23) combined Seurat with URD to delineate the anchoring of developmental trajectories to their spatial origins. Monocle2 has also been applied to reconstruct the pseudospace of single-cell transcription, and combining it with a pooled single-cell genetic screen enabled McFaline-Figueroa et al. (65) to identify continuous waves of gene regulation in the inner and outer cells during the epithelial-to-mesenchymal transition. The histo-cytometric multidimensional analysis pipeline (CytoMAP) includes a pseudospace function for revealing the spatial relationships among tumor microenvironments (99), and pseudospatial relationships based on Bayesian Gaussian process latent variable modeling (BGPLVM) reveal trajectories of tissue migration and adaptation (66). Pseudospace can help investigate how cells are distributed across different tissue compartments and reveal the fundamental relationships among cells. However, the scope of application of the current pseudospace methods is limited, and the accuracy needs to be further benchmarked. Most methods identify the pseudospace ordering of cells based on known spatial marker genes, which can only be used for simple pseudospace trajectories reconstruction, so many challenges remain in spatial analysis using computational mapping. Therefore, scRNA-seq defines individual cell types and also resolves the spatial architecture, providing information on the lineage relations, the regulatory circuitry, and the interactions among them.

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

Although spatial localization could be inferred from scRNA-seq data in some contexts where comprehensive regional landmarks are available (as discussed above), this computational retrospective location mapping has several limitations and must be rigorously tested. For example, the availability of known spatial landmarks is currently insufficient, and there is a lack of temporal resolution for a developmental-stage-matched mapping. Hence, it is essential to resolve lineage status in its native tissue settings with direct spatial transcriptomic approaches in real time. Geo-seq employs a mapping pipeline that uses extracted zip code genes from the embryonic tissues to restore single cells from the stage-matched embryos to their positions with high accuracy (78). In this seminal work, 94% of single cells obtained from known positions of the embryo were successfully allocated to their source address. More importantly, the in vitro cultured stem cells can be mapped to the embryonic locations to characterize the prospective cell fates as well, thus demonstrating the utility of zip code mapping for revealing the properties of stem cells (29, 69, 76).

In addition, the methodologies for extraction of the essential spatial coordinate genes are still in their infancy (21, 77). These issues need to be further addressed, and innovative research and development in computer algorithms are expected (72, 104). A machine learning protocol that uses analytical probing of large spatial transcriptome data sets could identify the hidden layer of spatial information much more efficiently.

SINGLE-CELL AND SPATIAL TRANSCRIPTOME APPROACHES TO MAP CELL FATE ALONG DEVELOPMENTAL TRAJECTORIES IN GASTRULATION: THE ENDODERMAL LINEAGE

The development of the mammalian embryo features a regulative mode of cell differentiation and tissue patterning, making it amenable to the quantitative analysis of cell fate decisions at granular temporal and spatial resolution. In contrast to preimplantation embryos, in which only a small number of cells are undergoing pluripotency transition, the gastrulating embryo increases its cell numbers from approximately 600 to more than 15,000 during a 24-hour period of development and regionalizes the progenitors of different lineages to the specific domains in three germ layers (43). Although a fate map has been constructed for the gastrulating embryo (107), how lineage development in each cell population is coordinated is important for establishing the body plan. Recently, Peng et al. (79) obtained holistic information on spatiotemporal transcriptomic activity at the critical developmental milestone of gastrulation. The architecture of gene profiling provided a molecular basis for annotating spatial variance during lineage commitment in gastrulation. The developmental changes underpinning the segregation of multiple lineages and the establishment of the ectoderm, mesoderm, and endoderm during the gastrulation process, particularly regarding the spatial variances, lay the foundation for future organogenesis and stem cell–based translational medicine.

To dissect the molecular map of cellular differentiation from the onset of gastrulation to early organogenesis, Pijuan-Sala et al. (81) collected 116,312 single cells at consecutive postimplantation developmental stages using a 10x Genomics platform. Another study further extended the time window to cover organogenesis until embryonic day 13.5 (E13.5) based on single-cell combinatorial indexing and completed an atlas of approximately 2 million cells (13). These comprehensive landscapes achieve a profiling of lineage trajectories in vivo at single-cell resolution, highlight the concurrent canalization and plasticity of embryonic specification, and provide a framework to reconstruct complex developmental trees at unprecedented depths. While the lineage information for the three germ layers gleaned from single-cell studies corroborates prior knowledge about the delineation of tissue progenitors, the bifurcation and convergence of endodermal tissues represents new biological evidence that supplements results from conventional fate mapping studies.

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

From an evolutionary perspective, approximately 85% of described animal phyla are triploblastic, consisting of three definitive germ layers—an outer ectodermal germ layer, an inner endodermal germ layer, and a middle mesodermal germ layer—whereas diploblastic animals lack the middle mesodermal germ layer (98, 113). A feeding process model suggests that the endoderm originates much earlier than the other two germ layers (31), and the transcription program of the mesoderm is the last to be induced. Histological analyses and live imaging in mouse embryos also revealed that cells destined to be mesoderm form last—transiting from an epithelial state to a mesenchymal state, then ingressing the primitive streak and migrating into the space between the visceral endoderm and epiblast (34, 84). Endoderm cells emerge first at the blastocyst stage, when the primitive endoderm segregates from the inner cell mass as a squamous epithelium. The primitive endoderm further differentiates into two populations: the visceral endoderm, which constitutes the outer epithelial layer that surrounds the extraembryonic ectoderm and epiblast, and the parietal endoderm after an epithelial–mesenchymal transition (6, 26). Before gastrulation, the parietal endoderm and visceral endoderm cells are responsible for nutrient and waste exchange between the maternal tissues and the embryo. At the end of gastrulation, the definitive (embryonic) endoderm arises and contributes to the gut tube (74). Fate mapping and lineage analysis studies have shown that epiblast cells might continue to be recruited to the visceral endoderm prior to the onset of gastrulation (106), and presumptive definitive endoderm cells are recruited to the preexisting layer of visceral endoderm from the primitive streak via a transitory residence in the mesoderm during gastrulation (47).

A complication in endoderm lineage development is that extraembryonic endoderm cells share a significant level of similarity in signature genes with embryonic endoderm (92), which may indicate a common mechanism of lineage specification. Echoing this idea, comprehensive scRNA-seq analyses have identified the convergence of primitive streak–derived definitive endoderm with visceral endoderm–derived cells at the molecular level as well as the continuum of status shifting of endodermal cells (75, 81). The single-cell molecular recording of gut endoderm cells confirmed the convergent transcriptional regulation of extraembryonic and embryonic endoderm cells in forming the gut endoderm (14) (**Figure 1**). However, the spatial relationship between the origins of endoderm cells and the potential locality of the mesendoderm progenitor is unclear.

The integration of the spatial transcriptome of embryos across the developmental stages in gastrulation mouse embryos has provided molecular annotation for the genealogy of defined cell populations that may be functionally connected by the proximity of their transcriptomes and locations. This has also enabled the construction of the developmental trajectory of the epiblast to the germ layer derivatives and the underpinning transcriptional and signaling activity in real time and space (79). The spatiotemporal transcriptome from the inner cell mass to E7.5 further showed that the endoderm populations are transcriptionally distinctive from the epiblast and the mesoderm. In comparison with the classic view that the embryonic endoderm derives solely from the epiblast at the late gastrulation stages, the endoderm cell lineage is separated from the epiblast-derived ectoderm and mesoderm as early as E4.5. The extraembryonic endoderm and definitive endoderm share many regulatory networks. Strikingly, although descendants of extraembryonic endoderm constitute approximately 15% of the gut tube at midgestation (74, 75), most cells within spatial endoderm domains at the E7.5 stage were associated with the visceral endoderm at earlier stages (**Figure 1**). The spatiotemporal transcriptome also revealed a shorter distance between the distal mesoderm and the endoderm in the vicinity, suggesting the potential existence of a mesendoderm progenitor. Notably, the progressive predominance of the distal population over the proximal population may reflect either the replacement of the visceral endoderm by the recruited population of the definitive endoderm (51, 108) or the acquisition of the molecular attributes of the definitive endoderm by the proximal endoderm (47, 119), pointing to the position-defined lineage progression.

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

Moreover, signaling activity enrichment analysis suggested a novel role of Hippo/Yap in regulating the segregation of the extraembryonic and embryonic endoderm. These examples showcase the importance of a holistic view of developmental processes—in which understanding can be achieved only when many of the seemingly temporally and spatially distinct working parts are considered together—to complement a reductionist molecule- and cell-level understanding (79).

OUTLOOK

Single-cell analysis has entered a new phase—from characterizing cellular heterogeneity to interpreting the role of spatial organization. In addition, single-cell multiomics has recently emerged as a powerful technology through which different layers of genomic output—and hence cell identity and function—can be recorded simultaneously. Integrating various components of the epigenome into multiomics measurements allows researchers to study cellular heterogeneity at different timescales and to discover new layers of molecular connectivity between the genome and its functional output.

Early embryo development and pluripotent stem cells are ideal models for exploring the mechanisms underpinning lineage differentiation. The in vivo embryo uses regulation and canalization at multiple layers to safeguard the developmental process. The multiscale integration of modular networks of gene expression and signaling and their interactions in spatial and temporal contexts orchestrated with epigenetic cues constitute the core of developmental mechanisms. The emergence of innovative technologies has made it feasible to address critical questions about how a cell makes fate choices in different temporal and spatial contexts. Among the emerging technologies, single-cell analysis and spatial transcriptome profiling provide effective tools for characterizing the origin and trajectory of cell lineages during early development (79). Together with single-cell imaging tracing (61, 64) and knowledge of the positions of individual cells in the developing embryo, these technologies could provide comprehensive insights into the interplay among cell movement, signal transduction pathways, transcription factors, and the chromatin packaging of the genome for defining the lineage of every cell type.

Ethical and legal constraints on human embryo research mean that investigations of lineage determination during human early embryo development are conducted mainly in preimplantation embryos, though the sustainable development of human embryos in vitro may open up new opportunities (20, 89, 90, 124, 125). Importantly, the discovery that human and nonhuman primate embryos develop similarly (70) and strides in long-term in vitro culturing of monkey embryos beyond gastrulation indicate the feasibility of elucidating the lineage specification in postimplantation primate embryos by using scRNA-seq and spatial transcriptome analysis (59, 73). Eventually, comparative study of gastrulation in nonhuman primates and mice and correlations with in vitro and ex vivo experimental models (20, 30, 90) could unravel the molecular architecture and lineage specification of the developmental processes at gastrulation, a major milestone of embryonic development.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors would like to thank Lihui Lin and Li Wang for critical reading of the manuscript and providing invaluable comments. This work was supported in part by the National Key

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

Basic Research and Development Program of China (2018YFA0108000, 2018YFA0107200, 2017YFA0102700, 2015CB964500, and 2014CB964804), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16020501 and XDA16020404), the National Natural Science Foundation of China (31871456, 31661143042, 91519314, 31630043, 31571513, and 31430058), the Guangdong Basic and Applied Basic Research Foundation (2019B151502054), the Science and Technology Planning Project of Guangdong Province (2017B030314056), and the Frontier Research Program (2018GZR110105013) and Jiazi Research Innovative Project (198H011211) of the Guangzhou Regenerative Medicine and Health Guangdong Laboratory.

LITERATURE CITED

1. Achim K, Pettit JB, Saraiva LR, Gavriouchkina D, Larsson T, et al. 2015. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33:503–9
2. Achim K, Vergara HM, Pettit JB. 2018. Spatial transcriptomics: constructing a single-cell resolution transcriptome-wide expression atlas. *Methods Mol. Biol.* 1649:111–25
3. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, et al. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14:1083–86
4. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A. 2018. Whole-organism clone tracing using single-cell sequencing. *Nature* 556:108–12
5. Amodio M, van Dijk D, Srinivasan K, Chen WS, Mohsen H, et al. 2019. Exploring single-cell data with deep multitasking neural networks. *Nat. Methods* 16:1139–45
6. Antonica F, Orietti LC, Mort RL, Zernicka-Goetz M. 2019. Concerted cell divisions in embryonic visceral endoderm guide anterior visceral endoderm migration. *Dev. Biol.* 450:132–40
7. Arnold SJ, Robertson EJ. 2009. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* 10:91–103
8. Baron CS, van Oudenaarden A. 2019. Unravelling cellular relationships during development and regeneration using genetic lineage tracing. *Nat. Rev. Mol. Cell Biol.* 20:753–65
9. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, et al. 2018. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37:38–44
10. Birnbaum KD. 2018. Power in numbers: single-cell RNA-seq strategies to dissect complex tissues. *Annu. Rev. Genet.* 52:203–21
11. Callaway E. 2017. The trickiest family tree in biology. *Nature* 547:20–22
12. Cannoodt R, Saelens W, Sichien D, Tavernier S, Janssens S, et al. 2016. SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. bioRxiv 079509. **https://doi.org/10.1101/079509**
13. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566:496–502
14. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, et al. 2019. Molecular recording of mammalian embryogenesis. *Nature* 570:77–82
15. Chen J, Suo S, Tam PP, Han JJ, Peng G, Jing N. 2017. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nat. Protoc.* 12:566–80
16. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090
17. Christodoulou N, Kyprianou C, Weberling A, Wang R, Cui G, et al. 2018. Sequential formation and resolution of multiple rosettes drive embryo remodelling after implantation. *Nat. Cell Biol.* 20:1278–89
18. Crosetto N, Bienko M, van Oudenaarden A. 2015. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* 16:57–66
19. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, et al. 2018. A single-cell transcriptome atlas of the aging Drosophila brain. *Cell* 174:982–98.e20
20. Deglincerti A, Croft GF, Pietila LN, Zernicka-Goetz M, Siggia ED, Brivanlou AH. 2016. Self-organization of the in vitro attached human embryo. *Nature* 533:251–54

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

21. Dries R, Zhu Q, Eng C-HL, Sarkar A, Bao F, et al. 2019. Giotto, a pipeline for integrative analysis and visualization of single-cell spatial transcriptomic data. bioRxiv 701680. **https://doi.org/10.1101/701680**
22. Eng CL, Lawson M, Zhu Q, Dries R, Koulena N, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 568:235–39
23. Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 360:eaar3131
24. Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJ. 2013. Using networks to measure similarity between genes: association index selection. *Nat. Methods* 10:1169–76
25. Gawad C, Koh W, Quake SR. 2016. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* 17:175–88
26. Grapin-Botton A, Melton DA. 2000. Endoderm development: from patterning to organogenesis. *Trends Genet*. 16:124–30
27. Griffiths JA, Scialdone A, Marioni JC. 2018. Using single-cell genomics to understand developmental processes and cell fate decisions. *Mol. Syst. Biol.* 14:e8046
28. Haghverdi L, Buettner F, Theis FJ. 2015. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 31:2989–98
29. Han X, Luo S, Peng G, Lu JY, Cui G, et al. 2018. Mouse knockout models reveal largely dispensable but context-dependent functions of lncRNAs during development. *J. Mol. Cell Biol.* 10:175–78
30. Harrison SE, Sozen B, Christodoulou N, Kyprianou C, Zernicka-Goetz M. 2017. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science* 356:eaal1810
31. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I. 2015. Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* 519:219–22
32. Hoppe PS, Coutu DL, Schroeder T. 2014. Single-cell technologies sharpen up mammalian stem cell research. *Nat. Cell Biol.* 16:919–27
33. Iacono G, Massoni-Badosa R, Heyn H. 2019. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol*. 20:110
34. Ichikawa T, Nakazato K, Keller PJ, Kajiura-Kobayashi H, Stelzer EH, et al. 2013. Live imaging of whole mouse embryos during gastrulation: migration analyses of epiblast and mesodermal cells. *PLOS ONE* 8:e64506
35. Irie N, Kuratani S. 2014. The developmental hourglass model: a predictor of the basic body plan. *Development* 141:4649–55
36. Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, et al. 2018. Developmental barcoding of whole mouse via homing CRISPR. *Science* 361:eaat9804
37. Kalinka AT, Tomancak P. 2012. The evolution of early animal embryos: conservation or divergence. *Trends Ecol. Evol.* 27:385–93
38. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, et al. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468:811–14
39. Kalisky T, Blainey P, Quake SR. 2011. Genomic analysis at the single-cell level. *Annu. Rev. Genet.* 45:431–45
40. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, et al. 2017. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358:194–99
41. Kelsey G, Stegle O, Reik W. 2017. Single-cell epigenomics: recording the past and predicting the future. *Science* 358:69–75
42. Kester L, van Oudenaarden A. 2018. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* 23:166–79
43. Kojima Y, Tam OH, Tam PP. 2014. Timing of developmental events in the early mouse embryo. *Semin. Cell Dev. Biol.* 34:65–75
44. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. 2015. The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58:610–20
45. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, et al. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16:1289–96

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

46. Kretzschmar K, Watt FM. 2012. Lineage tracing. *Cell* 148:33–45
47. Kwon GS, Viotti M, Hadjantonakis AK. 2008. The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. *Dev. Cell* 15:509–20
48. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, et al. 2018. RNA velocity of single cells. *Nature* 560:494–98
49. Lafzi A, Moutinho C, Picelli S, Heyn H. 2018. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat. Protoc.* 13:2742–57
50. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform*. 9:559
51. Lawson KA, Pedersen RA. 1987. Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. *Development* 101:627–52
52. Lein E, Borm LE, Linnarsson S. 2017. The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* 358:64–69
53. Levin M, Anavy L, Cole AG, Winter E, Mostov N, et al. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature* 531:637–41
54. Li M, Izpisua Belmonte JC. 2018. Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.* 20:382–92
55. Lin C, Jain S, Kim H, Bar-Joseph Z. 2017. Using neural networks for reducing the dimensions of single-cell RNA-Seq data. *Nucleic Acids Res*. 45:e156
56. Liu D, Xu C, He W, Xu Z, Fu W, et al. 2019. AutoGenome: an AutoML tool for genomic research. bioRxiv 842526. **https://doi.org/10.1101/842526**
57. Lu R, Neff NF, Quake SR, Weissman IL. 2011. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* 29:928–33
58. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. 2014. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* 11:360–61
59. Ma H, Zhai J, Wan H, Jiang X, Wang X, et al. 2019. In vitro culture of cynomolgus monkey embryos beyond early gastrulation. *Science* 366:seaax7890
60. Marioni JC, Arendt D. 2017. How single-cell genomics is changing evolutionary and developmental biology. *Annu. Rev. Cell Dev. Biol.* 33:537–53
61. Massarwa R, Niswander L. 2013. In toto live imaging of mouse morphogenesis and new insights into neural tube closure. *Development* 140:226–36
62. Maynard KR, Jaffe AE, Martinowich K. 2019. Spatial transcriptomics: putting genome-wide expression on the map. *Neuropsychopharmacology* 45:232–33
63. Mazzarello P. 1999. A unifying concept: the history of cell theory. *Nat. Cell Biol.* 1:E13–15
64. McDole K, Guignard L, Amat F, Berger A, Malandain G, et al. 2018. In toto imaging and reconstruction of post-implantation mouse development at the single-cell level. *Cell* 175:859–76.e33
65. McFaline-Figueroa JL, Hill AJ, Qiu X, Jackson D, Shendure J, Trapnell C. 2019. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nat. Genet.* 51:1389–98
66. Miragaia RJ, Gomes T, Chomka A, Jardine L, Riedel A, et al. 2019. Single-cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *Immunity* 50:493–504.e7
67. Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, et al. 2018. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362:eaau5324
68. Mooijman D, Dey SS, Boisset JC, Crosetto N, van Oudenaarden A. 2016. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* 34:852–56
69. Mulas C, Chia G, Jones KA, Hodgson AC, Stirparo GG, Nichols J. 2018. Oct4 regulates the embryonic axis and coordinates exit from pluripotency and germ layer specification in the mouse embryo. *Development* 145:dev159103
70. Nakamura T, Okamoto I, Sasaki K, Yabuta Y, Iwatani C, et al. 2016. A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* 537:57–62

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

71. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. 2011. Tumour evolution inferred by single-cell sequencing. *Nature* 472:90–94
72. Nitzan M, Karaiskos N, Friedman N, Rajewsky N. 2019. Gene expression cartography. *Nature* 576:132–37
73. Niu Y, Sun N, Li C, Lei Y, Huang Z, et al. 2019. Dissecting primate early post-implantation development using long-term in vitro embryo culture. *Science* 366:eaaw5754
74. Nowotschin S, Hadjantonakis AK, Campbell K. 2019. The endoderm: a divergent cell lineage with many commonalities. *Development* 146:dev150920
75. Nowotschin S, Setty M, Kuo YY, Liu V, Garg V, et al. 2019. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* 569:361–67
76. Osteil P, Studdert JB, Goh HN, Wilkie EE, Fan X, et al. 2019. Dynamics of Wnt activity on the acquisition of ectoderm potency in epiblast stem cells. *Development* 146:dev172858
77. Park J, Choi W, Tiesmeyer S, Long B, Borm LE, et al. 2019. Segmentation-free inference of cell types from in situ transcriptomics data. bioRxiv 800748. **https://doi.org/10.1101/800748**
78. Peng G, Suo S, Chen J, Chen W, Liu C, et al. 2016. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell* 36:681–97
79. Peng G, Suo S, Cui G, Yu F, Wang R, et al. 2019. Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* 572:528–32
80. Picelli S. 2017. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biol.* 14:637–50
81. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, et al. 2019. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566:490–95
82. Proserpio V, Lönnberg T. 2016. Single-cell technologies are revolutionizing the approach to rare cells. *Immunol. Cell Biol.* 94:225–29
83. Quinlan GA, Williams EA, Tan SS, Tam PP. 1995. Neuroectodermal fate of epiblast cells in the distal region of the mouse egg cylinder: implication for body plan organization during early embryogenesis. *Development* 121:87–98
84. Rivera-Pérez JA, Hadjantonakis AK. 2014. The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harb. Perspect. Biol.* 7:a015867
85. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, et al. 2019. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363:1463–67
86. Saelens W, Cannoodt R, Todorov H, Saeys Y. 2019. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* 37:547–54
87. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33:495–502
88. Scialdone A, Tanaka Y, Jawaid W, Moignard V, Wilson NK, et al. 2016. Resolving early mesoderm diversification through single-cell expression profiling. *Nature* 535:289–93
89. Shahbazi MN, Jedrusik A, Vuoristo S, Recher G, Hupalowska A, et al. 2016. Self-organization of the human embryo in the absence of maternal tissues. *Nat. Cell Biol.* 18:700–8
90. Shahbazi MN, Zernicka-Goetz M. 2018. Deconstructing and reconstructing the mouse and human early embryo. *Nat. Cell Biol.* 20:878–87
91. Shapiro E, Biezuner T, Linnarsson S. 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14:618–30
92. Sherwood RI, Jitianu C, Cleaver O, Shaywitz DA, Lamenzo JO, et al. 2007. Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Dev. Biol.* 304:541–55
93. Solnica-Krezel L. 2005. Conserved patterns of cell movements during vertebrate gastrulation. *Curr. Biol.* 15:R213–28
94. Solnica-Krezel L, Sepich DS. 2012. Gastrulation: making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* 28:687–717
95. Sozen B, Amadei G, Cox A, Wang R, Na E, et al. 2018. Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat. Cell Biol.* 20:979–89
96. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, et al. 2018. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* 36:469–73

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

97. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353:78–82
98. Steinmetz PRH, Aman A, Kraus JEM, Technau U. 2017. Gut-like ectodermal tissue in a sea anemone challenges germ layer homology. *Nat. Ecol. Evol.* 1:1535–42
99. Stoltzfus CR, Filipek J, Gern BH, Olin BE, Leal JM, et al. 2019. CytoMAP: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. bioRxiv 769877. **https://doi.org/10.1101/769877**
100. Street K, Risso D, Fletcher RB, Das D, Ngai J, et al. 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom*. 19:477
101. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, et al. 2019. Comprehensive integration of single-cell data. *Cell* 177:1888–902.e21
102. Stuart T, Satija R. 2019. Integrative single-cell analysis. *Nat. Rev. Genet.* 20:257–72
103. Suo S, Zhu Q, Saadatpour A, Fei L, Guo G, Yuan GC. 2018. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep*. 25:1436–45.e3
104. Svensson V, Teichmann SA, Stegle O. 2018. SpatialDE: identification of spatially variable genes. *Nat. Methods* 15:343–46
105. Talwar D, Mongia A, Sengupta D, Majumdar A. 2018. AutoImpute: autoencoder based imputation of single-cell RNA-seq data. *Sci. Rep.* 8:16329
106. Tam PP, Beddington RS. 1992. Establishment and organization of germ layers in the gastrulating mouse embryo. *Ciba Found. Symp.* 165:27–41
107. Tam PP, Behringer RR. 1997. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* 68:3–25
108. Tam PP, Khoo PL, Wong N, Tsang TE, Behringer RR. 2004. Regionalization of cell fates and cell movement in the endoderm of the mouse gastrula and the impact of loss of *Lhx1*(*Lim1*) function. *Dev. Biol.* 274:171–87
109. Tam PP, Loebel DA. 2007. Gene function in mouse embryogenesis: Get set for gastrulation. *Nat. Rev. Genet.* 8:368–81
110. Tam PP, Quinlan GA. 1996. Mapping vertebrate embryos. *Curr. Biol.* 6:104–6
111. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6:377–82
112. Tatapudy S, Aloisio F, Barber D, Nystul T. 2017. Cell fate decisions: emerging roles for metabolic signals and cell morphology. *EMBO Rep*. 18:2105–18
113. Technau U, Scholz CB. 2003. Origin and evolution of endoderm and mesoderm. *Int. J. Dev. Biol.* 47:531–39
114. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, et al. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32:381–86
115. Tritschler S, Büttner M, Fischer DS, Lange M, Bergen V, et al. 2019. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* 146:dev170506
116. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, et al. 2017. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* 14:935–36
117. van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9:2579–605
118. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, et al. 2019. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* 16:987–90
119. Viotti M, Nowotschin S, Hadjantonakis AK. 2014. SOX17 links gut endoderm morphogenesis and germ layer segregation. *Nat. Cell Biol.* 16:1146–56
120. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360:981–87
121. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, et al. 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol*. 20:59
122. Wu AR, Wang J, Streets AM, Huang Y. 2017. Single-cell transcriptional analysis. *Annu. Rev. Anal. Chem.* 10:439–62

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

123. Zhang J, Zhao J, Dahan P, Lu V, Zhang C, et al. 2018. Metabolism in pluripotent stem cells and early mammalian development. *Cell Metab*. 27:332–38
124. Zheng Y, Xue X, Shao Y, Wang S, Esfahani SN, et al. 2019. Controlled modelling of human epiblast and amnion development using stem cells. *Nature* 573:421–25
125. Zhou F, Wang R, Yuan P, Ren Y, Mao Y, et al. 2019. Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature* 572:660–64

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

Annual Review of Genomics and Human Genetics

Volume 21, 2020

Contents

The Long Journey from Diagnosis to Therapy
Kay E. Davies 1

An Accidental Genetic Epidemiologist
Robert C. Elston 15

Enhancer Predictions and Genome-Wide Regulatory Circuits
Michael A. Beer, Dustin Shigaki, and Danwei Huangfu 37

Progress, Challenges, and Surprises in Annotating the Human Genome
Daniel R. Zerbino, Adam Frankish, and Paul Flicek 55

RNA Conformation Capture by Proximity Ligation
Grzegorz Kudla, Yue Wan, and Aleksandra Helwak 81

Cell Lineage Tracing and Cellular Diversity in Humans
Alexej Abyzov and Flora M. Vaccarino 101

Cultivating DNA Sequencing Technology After the Human Genome Project
Jeffery A. Schloss, Richard A. Gibbs, Vinod B. Makhijani, and Andre Marziali 117

Pangenome Graphs
Jordan M. Eizenga, Adam M. Novak, Jonas A. Sibbesen, Simon Heumos, Ali Ghaffaari, Glenn Hickey, Xian Chang, Josiah D. Seaman, Robin Rounthwaite, Jana Ebler, Mikko Rautiainen, Shilpa Garg, Benedict Paten, Tobias Marschall, Jouni Sirén, and Erik Garrison 139

Using Single-Cell and Spatial Transcriptomes to Understand Stem Cell Lineage Specification During Early Embryo Development
Guangdun Peng, Guizhong Cui, Jincan Ke, and Naihe Jing 163

The Genomics and Genetics of Oxygen Homeostasis
Gregg L. Semenza 183

The Genetics of Epilepsy
Piero Perucca, Melanie Bahlo, and Samuel F. Berkovic 205

Twenty-Five Years of Spinal Muscular Atrophy Research: From Phenotype to Genotype to Therapy, and What Comes Next
Brunhilde Wirth, Mert Karakaya, Min Jeong Kye, and Natalia Mendoza-Ferreira 231

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.

The Laminopathies and the Insights They Provide into the Structural and Functional Organization of the Nucleus
Xianrong Wong and Colin L. Stewart 263

Recent Advances in Understanding the Genetic Architecture of Autism
Caroline M. Dias and Christopher A. Walsh 289

Genomic Data Sharing for Novel Mendelian Disease Gene Discovery: The Matchmaker Exchange
Danielle R. Azzariti and Ada Hamosh 305

Genomically Aided Diagnosis of Severe Developmental Disorders
David R. FitzPatrick and Helen V. Firth 327

New Diagnostic Approaches for Undiagnosed Rare Genetic Diseases
Taila Hartley, Gabrielle Lemire, Kristin D. Kernohan, Heather E. Howley, David R. Adams, and Kym M. Boycott 351

Population Screening for Inherited Predisposition to Breast and Ovarian Cancer
Ranjit Manchanda, Sari Lieberman, Faiza Gaba, Amnon Lahad, and Ephrat Levy-Lahad 373

Genetic Influences on Disease Subtypes
Andy Dahl and Noah Zaitlen 413

How Natural Genetic Variation Shapes Behavior
Natalie Niepoth and Andres Bendesky 437

Credit for and Control of Research Outputs in Genomic Citizen Science
Christi J. Guerrini and Jorge L. Contreras 465

Looking Beyond GINA: Policy Approaches to Address Genetic Discrimination
Yann Joly, Charles Dupras, Miriam Pinkesz, Stacey A. Tovino, and Mark A. Rothstein 491

Models of Technology Transfer for Genome-Editing Technologies
Gregory D. Graff and Jacob S. Sherkow 509

Pedigrees and Perpetrators: Uses of DNA and Genealogy in Forensic Investigations
Sara H. Katsanis 535

The Regulation of Mitochondrial Replacement Techniques Around the World
I. Glenn Cohen, Eli Y. Adashi, Sara Gerke, César Palacios-González, and Vardit Ravitsky 565

Annu. Rev. Genom. Hum. Genet. 2020.21:163-181. Downloaded from www.annualreviews.org
Access provided by 65.21.193.86 on 02/24/22. For personal use only.