



Review

Behavioral Neuroscience in the Era of Genomics: Tools and Lessons for Analyzing High-Dimensional Datasets

Assa Bentzur ^{1,2}, Shahar Alon ^{2,*} and Galit Shohat-Ophir ^{1,*}

¹ The Mina & Everard Goodman Faculty of Life Sciences, Gonda Multidisciplinary Brain Research Center and Institute of Nanotechnology, Bar-Ilan University, Ramat Gan 5290002, Israel; abentzur@gmail.com

² The Alexander Kofkin Faculty of Engineering, Gonda Multidisciplinary Brain Research Center, Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat Gan 5290002, Israel

* Correspondence: shahar.alon@biu.ac.il (S.A.); galit.ophir@biu.ac.il (G.S.-O.)

† Lead contact.

Abstract: Behavioral neuroscience underwent a technology-driven revolution with the emergence of machine-vision and machine-learning technologies. These technological advances facilitated the generation of high-resolution, high-throughput capture and analysis of complex behaviors. Therefore, behavioral neuroscience is becoming a data-rich field. While behavioral researchers use advanced computational tools to analyze the resulting datasets, the search for robust and standardized analysis tools is still ongoing. At the same time, the field of genomics exploded with a plethora of technologies which enabled the generation of massive datasets. This growth of genomics data drove the emergence of powerful computational approaches to analyze these data. Here, we discuss the composition of a large behavioral dataset, and the differences and similarities between behavioral and genomics data. We then give examples of genomics-related tools that might be of use for behavioral analysis and discuss concepts that might emerge when considering the two fields together.

Keywords: behavioral analysis; large datasets; variance

Citation: Bentzur, A.; Alon, S.; Shohat-Ophir, G. Behavioral Neuroscience in the Era of Genomics: Tools and Lessons for Analyzing High-Dimensional Datasets. *Int. J. Mol. Sci.* **2022**, *23*, 3811. <https://doi.org/10.3390/ijms23073811>

Academic Editor: Changjong Moon

Received: 10 February 2022

Accepted: 29 March 2022

Published: 30 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

During the last decade, behavioral neuroscience has gone through a technology-driven revolution with the emergence of machine-vision and machine-learning algorithms. Researchers can now obtain unprecedented high-resolution and high-throughput recordings, tracking and analysis of freely behaving animals. This gives rise to large-scale datasets containing continuous measurements over time of hundreds of behavioral parameters (features) per animal, between pairs of interacting animals, and parameters that characterize emergent group behavior. For example, for many years researchers utilized behavioral features that were based on recorded track paths (using center of mass), and discarded the rest of the video data because storage was constrained by the available hardware. Recent advances in hardware technology can accommodate large data files, and new machine-vision algorithms allow for the tracking of body parts, postures [1,2] and even facial expressions in high resolution [3], giving rise to rich behavioral datasets. For instance, in a behavioral experiment lasting 15 min (with 30 frames per second), up to hundreds of features per animal per frame can be generated, resulting in gigabytes of data per experiment [4–17]. Analyzing such large and complex datasets, researchers use advanced computational approaches (reviewed by [18,19]). However, recent advances in computational approaches in other fields of biology that regularly deal with large datasets, such as genomics, might be beneficial as well. While several fields in biology underwent technological advancement around the same time [20–22], standardized protocols and tools for genomic and proteomic analysis developed in a faster manner. This is partially due to the universality of their measured outputs (nucleotides and peptides) and

their massive use in many biological systems, unlike tools for behavioral analysis that were developed separately for specific behavioral paradigms in each type of animal.

In this paper, we discuss the possibility of adapting approaches from genomics to behavioral studies. First, we introduce the challenges with analyzing behavioral data, and discuss the similarities and differences between datasets in the fields of genomics and behavior, and then provide examples of approaches from genomics that might fit behavioral datasets.

2. Social Behavior Generates High-Dimensional Datasets

Animals execute many motivated behaviors over the course of their lives, with the goal of surviving and reproducing. When these behaviors involve interacting with other members of the same species, they are considered social interactions [23]. In each individual, sensory systems continuously perceive signals from the environment, generating an internal representation of the outside world. This sensory information is integrated with the internal state of the animal, determining how readily sensory stimuli trigger the type and intensity of motor actions [24]. In a social environment, the behavior of individuals is further influenced by and affects the behavior of others, resulting in a highly dynamic environment, where each interaction can change the social context of subsequent interactions, leading to a variety of behavioral outcomes from what seems to be identical starting conditions [7,25–29]. Therefore, social interaction in groups is a dynamic and multi-dimensional phenomenon that is continuously shaped by neuronal processing of external information in each individual. This internal representation determines individual behavioral choices that in turn affect other individuals. The complex nature of this environment imposes conceptual challenges in the quantification and analysis of group behavior.

While social interactions in groups give rise to emergent structures that are beyond the sum of the individuals that constitute them, capturing and analyzing such phenomena starts from simple building blocks that are based on the size, orientation, and location of single individuals per frame of the recording (Figure 1). This information is extracted from movies using machine-vision algorithms that track individual animals over time. This seemingly basic information provides a rich repertoire of calculated features such as velocities, angles, changes in velocities, angular velocity, and distances from the center and edge of the arena. Extending these features to account for relative measurements between pairs of animals dramatically increases the number of features per frame, resulting in a rich dataset that can be used to construct social networks and to automatically classify behaviors of single animals and ones that involve several individuals, using machine-learning algorithms (Figure 1). Calculating the duration and frequency of behaviors, number of interactions, length of interactions and types of interactions can give rise to hundreds of features that represent distinct behavioral aspects in each experiment [5,25,30,31] (Figure 1). When individuals interact in a social environment such as a group, the dynamical nature of this environment significantly affects the behavior of individuals that constitute it. One aspect of this is new types of behavioral parameters which are not exhibited when animals are alone, greatly enriching the behavioral dataset (Figure 1). Another aspect of this is that interactions between individuals constantly change their behavior, making them behave differently from how they would when alone, resulting in a somewhat interdependent dataset. Using robust tracking algorithms that maintain the identity of individuals over time, it is also possible to construct social networks that represent an emergent structure from the time-dependent interaction dataset [32,33].

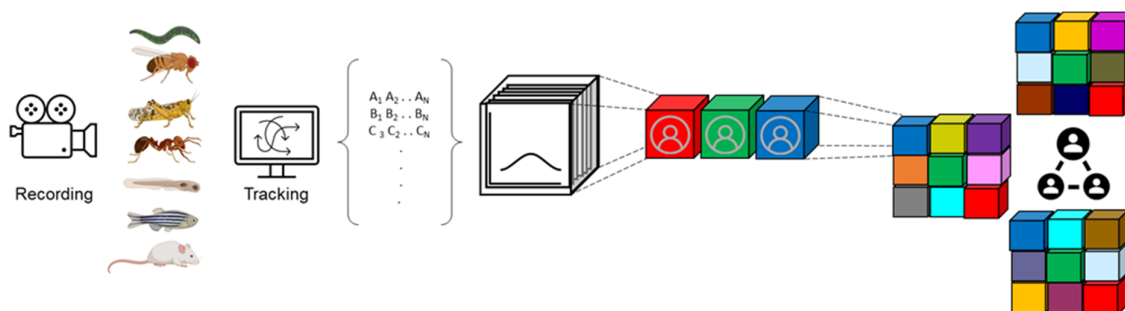


Figure 1. Quantification of behavior generates high-dimensional datasets. Behaviors are recorded over time, the resulting movies are then used to track the position, size, and orientation of all animals in each frame. Tracking data are used to compute various behavioral parameters for each individual per frame, which are then utilized to construct a distribution of behavior between individuals. Finally, interactions between individuals are quantified and are used to generate social network structures (right).

3. Both Behavioral and Genomic Datasets Are High-Dimensional

Both genomic and behavioral datasets describe high-dimensional systems, corresponding to collection of molecules in a given cell, repertoire of cells within a tissue, or individuals in a group. A common hallmark shared by all these systems is that they are composed of different organizational levels, each of which possess multiple dimensions. For instance, social groups are a collection of individuals, which are built from tissues that contain different repertoire of cells, containing various molecules, all of which work together to generate composite emergent phenomena. One can describe complex systems by representing their multiple dimensions as variables/features. For instance, a given cell can be described by the repertoire of expressed RNAs and proteins, the combination of which accounts for different cell states, or in the case of a single neuron, the composition of its ion channels determines its possible neuronal output [34]. Neuronal networks can be described by relative ratios of neuronal types and their possible connectivity, whereas behavior can be described by the types and extent of actions depicted by an individual, and social groups can be described by all possible interactions between individuals. In practice, one can use multiple dimensions to describe each organizational layer, giving rise to a high-dimensional space. The conceptual similarity when describing each organizational layer suggests one can implement tools that were originally developed to analyze the complexity of a certain organizational level, such as in single-cell RNA sequencing (scRNAseq) data, to analyze a different level, such as in social group interactions.

4. Technological Advances in Other Fields Generate Large Datasets

In the past decade, the fields of genomics and proteomics have led to an explosion of novel insights into the basic building blocks of life and how they are organized. The main drivers in these fields have been technological advancements, which allowed for the generation of massive amounts of molecular data. These technological advancements include for example, scRNAseq in genomics, and improved mass spectrometry analysis in proteomics. Moreover, recently many of these technologies have been applied in situ, generating information about molecules with their 2D or 3D localization [35–41]. Technological advancements at the molecular level are also evident in functional neuroscience, and include, for example, high-resolution electrophysiological-based and calcium-based recordings of neurons in living, behaving animals [42–47]. These advancements increase the resolution and number of molecular or functional elements that are measured in a single sample, leading to an explosion in the amount of data generated from a single experiment [48]. For example, using expansion sequencing [35], ~2 terabytes of image data are generated per one tissue sample (~1 cm by ~1 cm in size).

5. Similarities and Differences between Genomics and Behavior Data

The generation of ‘big’ data in the aforementioned fields naturally brings about technical challenges. Both theoretical and computational approaches that fit large datasets are needed, and indeed, much progress has been made in this regard over the last few years [22,48–59]. Most of these approaches were developed for genomics, the field that generates the largest amount of data. For this reason, from this point onwards, we will focus on the field of genomics. However, before we introduce some of the computational approaches in genomics/transcriptomics and their possible relevance for behavioral neuroscience, we want to point out a clear difference between the data generated in these two different fields: ‘snapshots’ versus continuous measurement. Data acquisition in transcriptomics is composed of ‘snapshots’ in time, i.e., each dataset represents the expression of all accessible RNA in the sample at a particular time. In most cases, only one ‘snapshot’ is possible per sample; for example, one piece of hippocampal tissue can be thoroughly studied via single-cell sequencing or MERFISH [33], as the tissue is either dissolved or fixed, respectively, in the process. In contrast, in behavioral experiments, data acquisition is continuous, i.e., each dataset represents a value of all behavioral parameters per frame, over the entire duration of the experiment.

The ‘snapshot’ nature of the transcriptomics data leads to questions which focus on the relationship between the examined features (i.e., molecules) within a specific time point. This includes characterizing molecules in cells, defining the cell types present in a sample, and refining the definition of cell types into cell states. These aggregated entities (cell types and cell states) are easier to compare between samples from different ‘snapshots’. In contrast, continuous measurement in behavioral experiments leads to questions which focus on the dynamics, activity and other behavioral aspects over time, and in-group behavior analysis is focused on the emergent social structure over time [60], which can be compared between groups or conditions. Interestingly, new approaches in genomics/transcriptomics allow one to estimate trajectories of genes, cells and circuits from ‘snapshots’ in an attempt to enable a time-dependent analysis (for example, cell-type trajectories [61] and RNA velocity [62,63]).

In addition to the differences with respect to the time domain, two other differences might seem to create a gap between genomics data and behavioral data: the first is the number of measured variables (i.e., number of features), and the second is the dependence between variables. Regarding the number of features, behavioral data are usually characterized by tens to hundreds of features, a number limited by our ability to distinguish unique behavioral characteristics. In contrast, in genomics, thousands or even tens of thousands of features are typically measured, as the dataset scales with the number of measurable genes. As a result, the resulting behavioral datasets might seem to be significantly smaller compared to their genomics equivalents. However, the difference in the number of features is less dramatic when one considers the fact that in behavioral data, each feature is measured in every time frame, multiplying the size of the dataset by the number of frames in the experiment. Regarding the dependence between variables, one might think that measuring many variables in behavioral assays is fundamentally different from measuring gene expression in genomics, because the behavioral variables are more likely to be dependent. However, genes are also highly correlated in their expression and tend to work in concert [64–66]. In fact, if the expression of different genes was not strongly correlated, the high dimensionality of the problem would have made it impossible to analyze gene expression in single cells [58]. With the differences and similarities between behavioral and genomics dataset in mind, what kind of approaches can be ‘borrowed’ from genomics? Below are a few examples.

6. Dimensionality Reduction

The general idea of dimensionality reduction is to reduce the number of variables without creating a significant loss of information. In genomics, various methods to reduce

high-dimensional data into lower dimensional space are routinely used (reviewed by [67]), facilitating easier comparisons between conditions (Figure 2A). Fundamental methods for dimension reduction include principal component analysis (PCA), which finds orthogonal features of maximum variation; independent component analysis (ICA), which finds statistically independent features that best reconstruct the original data (Figure 2B); and nonnegative matrix factorization (NMF), which finds gene modules that combine expression across multiple correlated genes (Figure 2C). In addition to dimension reduction, new methods allow for the visualization of high-dimension data as 3D and even 2D plots that are easier to understand (albeit with reduced ability to perform formal analysis on the resulting plots) (Figure 2A right). These visualization methods include t-distributed stochastic neighbor embedding (t-SNE) [68]; methods based on k-nearest neighbor (KNN) graphs, which can be visualized according to a force-directed layout [69,70]; and the uniform manifold approximation and projection (UMAP) algorithm [71,72]. We note that in behavioral analysis, averaging behavioral variables over time and over individuals is commonly used. This approach can be thought of as a basic type of dimensionality reduction, as a large dataset is replaced with a simpler one in which there is a reduced dependence between the variables and detecting changes between variables is easier. However, adopting formal dimensionality-reduction methods in behavioral data can be informative and can help reveal underlying mechanisms which are not obvious when looking at means.

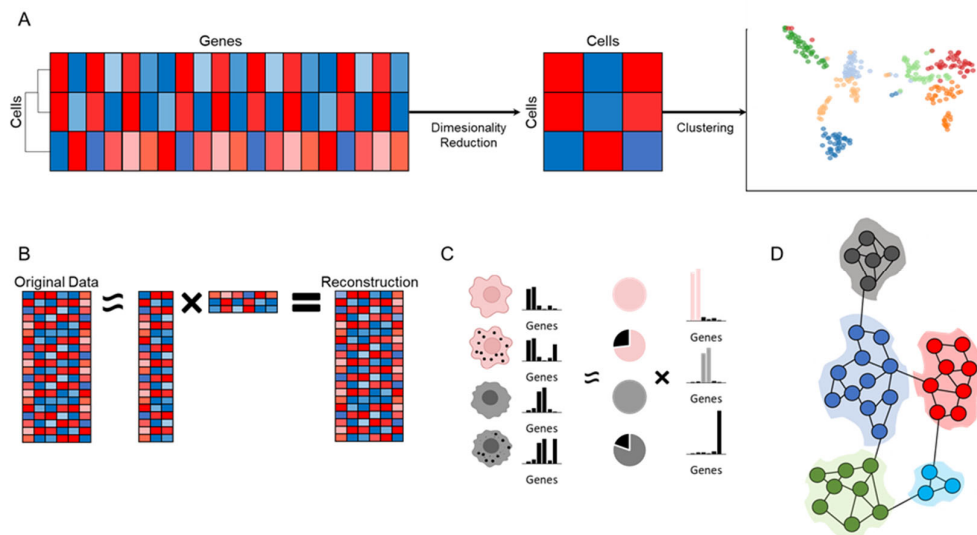


Figure 2. Dimensionality-reduction, clustering and matrix-decomposition approaches used in genomics to analyze expression patterns in high-dimensional datasets. **(A)** Reducing high-dimensional data into lower dimensional space enables clustering of data and visualization of cell types (represented by the different colors). **(B)** PCA, ICA and NMF are all based on matrix decomposition of the original dataset, using a set of describing vectors (reviewed by [67]). In PCA, the describing vectors are the principal components, whereas in ICA these vectors are chosen to maximize their independence. **(C)** In NMF, these vectors are termed gene modules, and can describe either the cell types (pink and gray cells) or cell states, the latter illustrated as a specific gene expressed within the cells (black dots). Image inspired by Kotliar et al. [73]. **(D)** Network Community Detection creates a graph that represents cells (nodes) and cell–cell similarity metrics (edges), which is then used to identify densely connected regions as clusters (reviewed by [74]).

7. Clustering Analysis

In genomics, clustering of data is a powerful tool to gain insights into high-dimensional data [75] (reviewed by [76,77]). This approach allows conversion of the time-dependent expression of thousands of genes into tens of tangible expression ‘profiles’. In

many cases, the genes in a given profile share a similar function, elucidating the functional role of the genes involved. Moreover, genes for which no functional information was available have become accessible using this approach. It is easy to see how this approach can be generalized from the time-dependent expression of genes into other time-dependent features (for example in behavioral measurements). In a typical single-cell RNA-sequencing experiment performed nowadays, clustering analysis is performed on the level of the cells. The motivation is to find cells which are similar to one another, therefore converting thousands (or more) of individual cells into tens of ‘types’ of cells (Figure 2A). The higher the number of cell types (or cell states, which are finer scale variations in cells within a cell type), the more homogeneous the cells within each type or state [78–80]. The cell groups (i.e., cell types and states) can then be compared to one another or across different experimental conditions in a given cell type. Two popular types of clustering approaches are hierarchical clustering and network community detection [75,77]. Hierarchical clustering works by joining cells iteratively into groups based on similarity metrics. In network community detection, a graph is generated to represent cells (nodes) and cell–cell similarity metrics (edges), and then densely connected regions of nodes are identified as clusters (Figure 2D). Both methods share the need for a cell-to-cell similarity metric. In other words, these methods can detect groups of cells given a similarity metric between them, and therefore defining such a metric becomes the main challenge. This is performed by deciding which genes should be used to define the distance between the cells, a problem which echoes the dimensionality-reduction topic mentioned above.

Clustering analysis can also be useful to extract meaningful patterns in complex behavioral datasets, where instead of genes/cells, one clusters behavioral parameters and conditions/genetic manipulations [25]. For example, we used hierarchical clustering to compare behavioral signatures of male *Drosophila* flies under various conditions. The behavioral signature of each group contained more than 50 behavioral parameters, including velocities, angles, distance between flies, duration and frequency of behaviors as well as features that describe the formation of social networks. Extracting meaningful patterns from this rich dataset required normalization of the various features followed by their clustering according to conditions and features. This type of analysis allows one to discriminate between conditions or genetic manipulations and to identify groups which exhibit similar patterns. Moreover, clustering of behavioral features can illuminate subsets of clustered parameters that are co-regulated under a certain condition/genotype, providing valuable information about the hallmark of the studied behavior [25]. In addition, one can use visualization algorithms such as t-SNE (broadly used in single-cell RNAseq experiments), to view individuals in 2D across conditions/genotypes, such that every point in the graph represents the behavioral repertoire/signature of an individual. The 2D representation of individuals by types/conditions can be used to test whether populations are composed of individuals that are similar or different from one another, or whether a single population is composed of subgroups, as can be seen in Figure 3.

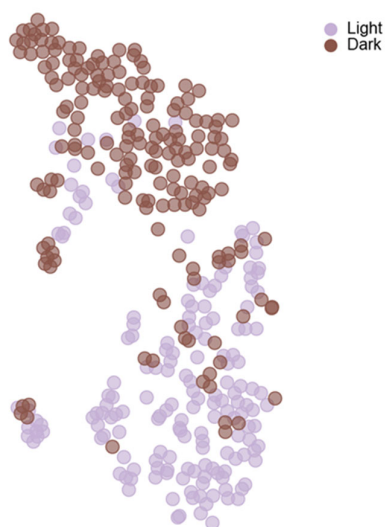


Figure 3. Methods used for visualization of single-cell transcriptomics can be used for behavioral data. t-SNE analysis of all behavioral parameters of individual wild-type flies. Each circle represents one individual. Flies were raised in groups for 3 days prior to testing in a 12 h light/dark cycle and were then tested as a group in FlyBowl arenas, either in light (purple) or in dark (brown) conditions. Behavioral recording and analysis were performed as in Bentzur et al. [25].

8. Variance as a Tool to Investigate Behavioral Phenomena

When trying to elucidate a biological mechanism, such as how the activity of specific genes affects a specific phenotype, we usually compare the means of experimental groups to controls, with the hope to record low variability between repeats for high statistical significance between the means of tested groups [81]. Behavioral data are notoriously variable, mainly because of inter-individual differences in voluntary actions and high sensitivity to even mild environmental changes, both of which pose a challenge to reproducing experimental results. Therefore, reducing behavioral variance is desirable in order to increase our ability to resolve the mean of an effect, mostly by increasing the number of repeats per experiment and by better control of test conditions [81].

Nevertheless, studying behavioral variance can have an added value, as differences between populations can exist in the distribution of the data (variance), without affecting averages [82]. Indeed, there are interesting examples for how variance is an additional and valuable feature to describe biological phenomena. For example, in the field of ecology, increased variance of certain parameters in a group is known to increase survivability of the group in a changing environment, a phenomenon known as bet hedging [83]. This means that actions taken by certain individuals are sub-optimal for their survival and reproduction in a constant environment but could be beneficial in an unpredictable environment. A classic example of this is the ratio between germinated and ungerminated plant seeds in a given year [83].

Applying this principle to social environments, which are inherently competitive and unpredictable, suggests that behavioral differences between individuals in a group can contribute to its overall success [84]. Interestingly, even in highly synchronized collective behaviors of schooling fish and swarming locusts, which are expected to have low variance between individuals, some level of interindividual variance is still required. Knebel et al. show that in locusts, inter-individual behavioral differences drive differences in inter-group differences, and that this can explain certain attributes of swarm formation [85]. Jolles et al. show that in fish, heterogeneous groups can form collective behaviors based on consistent inter-individual differences [10].

Social interactions between members in a group are a driving force for variance between individuals, since each individual develops its own unique interaction sequence

(i.e., with which group member, in what order, type of interaction, and its duration). Even assuming similar starting states for all individuals, the “trajectory” of each individual is actively determined by its interaction sequence, thus creating differences between members of the group [25]. Any given behavioral feature is represented by a series of values over time, per individual. One can extract from this raw data a distribution of values over time between all individuals (inter-individual variance) in a group and between groups (inter-group variance). These measures are useful as additional dimensions to the behavioral dataset, and as a tool to understand the contribution of prior experience, group composition or other biological aspects to social group dynamics. For instance, groups composed of socially raised *Drosophila* males exhibit high inter-individual and high inter-group variation in various behavioral features compared to groups composed of male flies that were raised in social isolation [25]. This implies that social enrichment facilitates behavioral variability between groups, such that the range of possible group dynamics exhibited by different groups is larger when they are composed of individuals with prior social experience.

9. The Concept of Individuality and Group Identity

A well-documented phenomenon in behavioral ecology is that individuals exhibit certain characteristics that are maintained over time. These can include the amount of attraction or aversion to others, relative activity levels and other attributes that are maintained by specific individuals over the duration of the experiment, and are considered as persistent inter-individual differences [86–89]. For example, Stern et al. showed that in *C. elegans*, even isogenic animals in a similar environment can have consistent behavioral biases that differ from the population average, and that this is regulated by neuromodulatory mechanisms [86], suggesting that persistent inter-individual differences are a result of specific biological mechanisms. Recent advances in computational neuroscience for analyzing high-dimensional behavioral data allow us to resolve inter-individual differences and even specific personalities, which can account for part of the overall variance in a population [90]. It will be interesting to see whether it is possible to connect these two aspects by attempting to correlate certain personality types with differences in the same mechanisms that underlie consistent behavioral biases.

One can extend the idea of individuality to social groups, where each group exhibits consistent features maintained over time that emerge from interactions between members of the group, suggestive of distinct group identities. For instance, when analyzing variance between groups composed of *Drosophila* flies that were raised in isolation or in groups, socially experienced flies exhibit higher inter-group variance, suggesting that each group developed emergent characteristics that make it distinguishable from other groups with seemingly identical individuals [25]. A recent study investigated the factors that shape and maintain variation in group structure in *Drosophila melanogaster*. Focusing on the positions of individuals within the network as a feature that shows variability between groups, they discovered that genotypes of the individuals comprising the group, their prior experience and environmental conditions contribute to inter-group variation [91]. Altogether, behavioral variation in group structure shares conceptual similarities with bet hedging. This raises interesting questions about the molecular and neuronal mechanisms that maintain behavioral persistence of both individuals and groups: does inter-group variance increase the survivability of groups in a changing environment, what is the effect of group size on survivability, and which neuromodulatory mechanisms control the magnitude of variance between groups?

10. Variance and Individuality in Genomics

An interesting case study for the possible learning opportunities from genomics in social behavior and vice versa is gene expression variance. Variance in expression per gene is tightly regulated in the level of tissues and organisms [92,93]. Changes in this regulation due to genetic mutations can create changes in morphological and physiological

traits and give rise to disease states [92,93]. Single-cell genomics allows us to measure the variance in gene expression between individual cells. However, explaining the source of this measured variance was elusive until recently. While it is known that stochasticity plays a major role in gene expression variance at the single-cell level, not enough was known about other possible sources of the measured variance. With the emergence of the field of spatially resolved transcriptomics, it is now possible to physically position the individual cells in tissues. This allows us to break down the variance in gene expression into additional measurable sources. In other words, how much of the variance in gene expression can be explained by cell type? How much of the remaining variance is explained by physical location in the tissue, or by cell–cell interactions? [94–99]. The analytic tools used to quantify the contributing factors to gene expression variance in single cells can potentially be of use for quantifying variance in the measurements of behavior in individuals in social settings [94–99]. The analogy here is that individuals can be akin to cells, leading to questions about what part of the behavioral variance is explained by certain parameters, such as distance between individuals, their interactions, etc. This analogy might also be useful in the other direction as well: is it time to start thinking about ‘individuality’ in cells? The prevailing way of thinking about cells is that they are characterized by their cell type; however, we are getting better and better at measuring cell states, and connecting them to physical factors, such as proximity to other cells, location in tissue, exposure to environmental factors such as hypoxia, and so on. These differences between individual cells of the same type, which are manifested in gene expression, can be persisted over time (which is not trivial to measure) or over space, as outlined above, and therefore can contribute to the concept of ‘individuality’ in cells. Evidence for this concept can be seen in a recent study demonstrating that variation in gene expression between cells of isogenic melanoma cells, that were previously considered as noise, are in fact persistent differences that are maintained over several cell divisions [100]. Lastly, Phillips et al. demonstrated that daughter cells in embryonic cell lineages maintain similar patterns of gene expression to their mother cells, providing a mechanistic explanation for the formation of inter-cell variation in gene expression patterns within tissues [101].

11. Future Perspective and Challenges

A hallmark of genomics research over the last two decades is data accessibility and standardization of analysis, an issue which is still lacking in the field of behavioral neuroscience research [102]. Ideally, this means that all raw data from each experiment are deposited in a way which is accessible to everyone and can be processed using standard tools which keep improving using community involvement. Software engineers are now routinely involved in the storage, processing and streamlining analysis of genomics data. As a result, constantly updated software is available for the community, for example the Seurat R toolkit [103,104] and Scanpy [105] for normalization, scaling, transformation and processing. Moreover, analytic tools allow for direct comparison between the analysis of multiple datasets which were generated by the same lab at different times, different labs, or even different modalities of measurements, without the need to re-analyze the raw data. Two examples of such an analysis are canonical correlation analysis (CCA), which aims to identify a set of variables that are maximally correlated between two datasets, and mutual nearest neighbors (MNNs), which detects variables mutually closest to each other across datasets [104].

Currently, all genomic-related studies are required to deposit the raw data at public archives, such as the Gene Expression Omnibus (GEO) database. This, in combination with standardized methods for sample preparation and analysis, facilitates further investigation of the original data by other research groups, and even allows for comparison between datasets from different sources. This is far from being the case for behavior datasets, due to various reasons. While the building blocks in genomics are strings of nucleic acids, it is hard to find a uniform basic element common to all behavioral measurements. It is even challenging to unify all features from a single experiment that contains velocities

(mm/sec), angle (rad), distances (mm), behaviors (% of frames) and durations of behaviors (sec), requiring normalization of all features using approaches such as Z-Scoring. Many labs do not even compare between experiments that were conducted on different days due to batch effects of day-to-day differences in behavior that result from mild variation in environmental conditions, nevermind of comparing animals with different genetic backgrounds. Adding to that is the use of various behavioral paradigms, arenas and lighting conditions that require adjusting tracking algorithms to each experimental setup. Lastly, behavioral neuroscience investigates a variety of animals, with different anatomical shapes, sizes and complex behaviors.

Although there are many challenges that impede standardization of behavioral data, some machine-vision algorithms such as CTRX [106] and DeepLabCut [107] are capable of tracking various organisms, providing hope that it will be possible at some point to compare between experiments and maybe model organisms using several basic features. While acquisition is achieved using different methods, we believe that it is possible to find universal parameters that can facilitate the standardization of the data. Firstly, these include general information about the experimental setup: type of animal, genotype, sex, age, and time of day is necessary. Secondly, technical specifications: illumination, frame rate, arena size and shape, temperature and humidity, number of animals and other parameters are crucial to create a comparable dataset. Lastly, parameters of the animal's coordinates, dimensions of its fitted shape and orientation per frame for the duration of the experiment. Depositing this raw data will facilitate analysis by other researchers, which also raises questions of how to optimize data storage when datasets are quickly growing. Maybe the most promising field in this respect is the study of social networks, which uses a set of features that describe individuals within a network and compares between different networks under various conditions. A step towards this direction was recently achieved in a study by Jezovit et al. [108], who compared the results of several studies that focused on social networks in *Drosophila melanogaster*. Although each study used a different approach to calculate network features, the authors documented similar effects of social isolation on network structure across all studies. Several studies attempted to compare social network structures between different organisms [109,110], suggesting that there are similarities in structures of groups from different species, although these datasets were generated using different techniques [111]. This further emphasizes the need to use a standardized set of basic behavioral features when depositing datasets. An in-depth discussion of the optimal set of features for this is instrumental in advancing the field of behavioral neuroscience.

To conclude, behavioral neuroscience is shifting towards large data. This poses several challenges and opportunities for advancing our understanding of phenomena which were once hard to analyze, such as behavioral variance. To advance the field, we can adopt approaches from other fields such as genomics and decide on protocols for best-practice dataset deposition, which will enable the development of analysis toolkits and other community resources.

Funding: This work was supported by the Israel Science Foundation Grants 174/19 and 384/14 granted to Galit Shohat-Ophir and by Israel Science Foundation Grants 2958/21 and 3363/21 granted to Shahar Alon.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cande, J.; Namiki, S.; Qiu, J.; Korff, W.; Card, G.M.; Shaevitz, J.W.; Stern, D.L.; Berman, G.J. Optogenetic dissection of descending behavioral control in *Drosophila*. *eLife* **2018**, *7*, e34275.
2. Roemschied, F.A.; Pacheco, D.A.; Ireland, E.C.; Li, X.; Aragon, M.J.; Pang, R.; Murthy, M. Flexible Circuit Mechanisms for Context-Dependent Song Sequencing. *bioRxiv* **2021**. <https://doi.org/10.1101/2021.11.01.466727>.
3. Dolensek, N.; Gehrlach, D.A.; Klein, A.S.; Gogolla, N. Facial expressions of emotion states and their neuronal correlates in mice. *Science* **2020**, *368*, 89–94.

4. Graving, J.M.; Chae, D.; Naik, H.; Li, L.; Koger, B.; Costelloe, B.R.; Couzin, I.D. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **2019**, *8*, e47994.
5. Robie, A.A.; Hirokawa, J.; Edwards, A.W.; Umayam, L.A.; Lee, A.; Phillips, M.L.; Card, G.M.; Korff, W.; Rubin, G.M.; Simpson, J.H.; et al. Mapping the Neural Substrates of Behavior. *Cell* **2017**, *170*, 393–406.e28.
6. Ariel, G.; Ayali, A. Locust Collective Motion and Its Modeling. *PLoS Comput. Biol.* **2015**, *11*, e1004522.
7. Shemesh, Y.; Sztainberg, Y.; Forkosh, O.; Shlapobersky, T.; Chen, A.; Schneidman, E. High-order social interactions in groups of mice. *eLife* **2013**, *2*, e00759.
8. Karamihalev, S.; Brivio, E.; Flachskamm, C.; Stoffel, R.; Schmidt, M.V.; Chen, A. Social dominance mediates behavioral adaptation to chronic stress in a sex-specific manner. *eLife* **2020**, *9*, e58723.
9. Elliott, E.; Ezra-Nevo, G.; Regev, L.; Neufeld-Cohen, A.; Chen, A. Resilience to social stress coincides with functional DNA methylation of the Crf gene in adult mice. *Nat. Neurosci.* **2010**, *13*, 1351–1353.
10. Jolles, J.W.; Boogert, N.J.; Sridhar, V.H.; Couzin, I.D.; Manica, A. Consistent Individual Differences Drive Collective Behavior and Group Functioning of Schooling Fish. *Curr. Biol.* **2017**, *27*, 2862–2868.e7.
11. Rosenthal, S.B.; Twomey, C.R.; Hartnett, A.T.; Wu, H.S.; Couzin, I.D. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 4690–4695.
12. Versace, E.; Caffini, M.; Werkhoven, Z.; De Bivort, B.L. Individual, but not population asymmetries, are modulated by social environment and genotype in *Drosophila melanogaster*. *Sci. Rep.* **2020**, *10*, 4480.
13. Rooke, R.; Rasool, A.; Schneider, J.; Levine, J.D. *Drosophila melanogaster* behaviour changes in different social environments based on group size and density. *Commun. Biol.* **2020**, *3*, 304.
14. Anpilov, S.; Shemesh, Y.; Eren, N.; Harony-Nicolas, H.; Benjamin, A.; Dine, J.; Oliveira, V.E.M.; Forkosh, O.; Karamihalev, S.; Hüttel, R.; et al. Wireless Optogenetic Stimulation of Oxytocin Neurons in a Semi-natural Setup Dynamically Elevates Both Pro-social and Agonistic Behaviors. *Neuron* **2020**, *107*, 644–655.e7.
15. Marshall, J.D.; Aldarondo, D.E.; Dunn, T.W.; Wang, W.L.; Berman, G.J.; Ölveczky, B.P. Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire. *Neuron* **2021**, *109*, 420–437.e8.
16. Netser, S.; Meyer, A.; Magalnik, H.; Zylbertal, A.; de la Zerdá, S.H.; Briller, M.; Bizer, A.; Grinevich, V.; Wagner, S. Distinct dynamics of social motivation drive differential social behavior in laboratory rat and mouse strains. *Nat. Commun.* **2020**, *11*, 5908.
17. Davidson, J.D.; Sosna, M.M.G.; Twomey, C.R.; Sridhar, V.H.; Leblanc, S.P.; Couzin, I.D. Collective detection based on visual information in animal groups. *J. R. Soc. Interface* **2021**, *18*, 20210142.
18. Smith, J.E.; Pinter-Wollman, N. Observing the unwatchable: Integrating automated sensing, naturalistic observations and animal social network analysis in the age of big data. *J. Anim. Ecol.* **2021**, *90*, 62–75.
19. Von Ziegler, L.; Sturman, O.; Bohacek, J. Big behavior: Challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacology* **2021**, *46*, 33–44.
20. Hong, G.; Lieber, C.M. Novel electrode technologies for neural recordings. *Nat. Rev. Neurosci.* **2019**, *20*, 330–345.
21. Hedlund, E.; Deng, Q. Single-cell RNA sequencing: Technical advancements and biological applications. *Mol. Asp. Med.* **2018**, *59*, 36–46.
22. Halder, A.; Verma, A.; Biswas, D.; Srivastava, S. Recent advances in mass-spectrometry based proteomics software, tools and databases. *Drug Discov. Today Technol.* **2021**, *39*, 69–79.
23. Wills, G.D.; Wesley, A.L.; Moore, F.R.; Sisemore, D.A. Social interactions among rodent conspecifics: A review of experimental paradigms. *Neurosci. Biobehav. Rev.* **1983**, *7*, 315–323.
24. Modi, M.; Shuai, Y.; Turner, G.C. The *Drosophila* Mushroom Body: From Architecture to Algorithm in a Learning Circuit. *Annu. Rev. Neurosci.* **2020**, *43*, 465–484.
25. Bentzur, A.; Ben-Shaanan, S.; Benichou, J.I.C.; Costi, E.; Levi, M.; Ilany, A.; Shohat-Ophir, G. Early Life Experience Shapes Male Behavior and Social Networks in *Drosophila*. *Curr. Biol.* **2021**, *31*, 670.
26. Burmeister, S.S.; Jarvis, E.D.; Fernald, R.D. Rapid Behavioral and Genomic Responses to Social Opportunity. *PLoS Biol.* **2005**, *3*, e363.
27. Karvat, G.; Kimchi, T. Acetylcholine Elevation Relieves Cognitive Rigidity and Social Deficiency in a Mouse Model of Autism. *Neuropsychopharmacology* **2014**, *39*, 831–840.
28. Weissbrod, A.; Shapiro, A.; Vasserman, G.; Edry, L.; Dayan, M.; Yitzhaky, A.; Hertzberg, L.; Feinerman, O.; Kimchi, T. Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nat. Commun.* **2013**, *4*, 2018.
29. Forkosh, O.; Karamihalev, S.; Roeh, S.; Alon, U.; Anpilov, S.; Touma, C.; Nussbaumer, M.; Flachskamm, C.; Kaplick, P.M.; Shemesh, Y.; et al. Identity domains capture individual differences from across the behavioral repertoire. *Nat. Neurosci.* **2019**, *22*, 2023–2028.
30. Kabra, M.; Robie, A.; Rivera-Alba, M.; Branson, S.; Branson, K. JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nat. Methods* **2013**, *10*, 64–67.
31. Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*; Bederson, B.B., Shneiderman, B., Eds.; Morgan Kaufmann: Burlington, MA, USA, 2003; pp. 364–371.
32. Croft, D.P.; James, R.; Krause, J. *Exploring Animal Social Networks*; Princeton University Press: Princeton, NJ, USA, 2008.
33. Whitehead, H. *Analyzing Animal Societies*; University of Chicago Press: Chicago, IL, USA, 2008.

34. Goillard, J.-M.; Marder, E. Ion Channel Degeneracy, Variability, and Covariation in Neuron and Circuit Resilience. *Annu. Rev. Neurosci.* **2021**, *44*, 335–357.
35. Keren, L.; Bosse, M.; Thompson, S.; Risom, T.; Vijayaragavan, K.; McCaffrey, E.; Marquez, D.; Angoshtari, R.; Greenwald, N.F.; Fienberg, H.; et al. MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Sci. Adv.* **2019**, *5*, eaax5851.
36. Moffitt, J.R.; Hao, J.; Wang, G.; Chen, K.H.; Babcock Hazen, P.; Zhuang, X. High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11046–11051.
37. Wang, X.; Allen, W.E.; Wright, M.A.; Sylwestrak, E.L.; Samusik, N.; Vesuna, S.; Evans, K.; Liu, C.; Ramakrishnan, C.; Liu, J.; et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **2018**, *361*, eaat5691.
38. Alon, S.; Goodwin, D.R.; Sinha, A.; Wassie, A.T.; Chen, F.; Daugharthy, E.R.; Bando, Y.; Kajita, A.; Xue, A.G.; Marrett, K.; et al. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **2021**, *371*, eaax2656.
39. Rodriques, S.G.; Stickels, R.R.; Goeva, A.; Martin, C.A.; Murray, E.; Vanderburg, C.R.; Welch, J.; Chen, L.M.; Chen, F.; Macosko, E.Z. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **2019**, *363*, 1463–1467.
40. Goltsev, Y.; Samusik, N.; Kennedy-Darling, J.; Bhate, S.; Hale, M.; Vazquez, G.; Black, S.; Nolan, G.P. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **2018**, *174*, 968–981.e15.
41. Ståhl, P.L.; Salmén, F.; Vickovic, S.; Lundmark, A.; Navarro, J.F.; Magnusson, J.; Giacomello, S.; Asp, M.; Westholm, J.O.; Huss, M.; et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **2016**, *353*, 78–82.
42. Driscoll, N.; Erickson, B.; Murphy, B.B.; Richardson, A.G.; Robbins, G.; Apollo, N.V.; Mentzelopoulos, G.; Mathis, T.; Hantanasirisakul, K.; Bagga, P.; et al. MXene-infused bioelectronic interfaces for multiscale electrophysiology and stimulation. *Sci. Transl. Med.* **2021**, *13*, eabf8629.
43. Liu, P.; Miller, E.W. Electrophysiology, Unplugged: Imaging Membrane Potential with Fluorescent Indicators. *Acc. Chem. Res.* **2020**, *53*, 11–19.
44. Kim, K.; Vöröslakos, M.; Seymour, J.P.; Wise, K.D.; Buzsáki, G.; Yoon, E. Artifact-free and high-temporal-resolution in vivo opto-electrophysiology with microLED optoelectrodes. *Nat. Commun.* **2020**, *11*, 2063.
45. Adam, Y. All-optical electrophysiology in behaving animals. *J. Neurosci. Methods* **2021**, *353*, 109101.
46. Tian, J.; Lin, Z.; Chen, Z.; Obaid, S.N.; Efimov, I.R.; Lu, L. Stretchable and Transparent Metal Nanowire Microelectrodes for Simultaneous Electrophysiology and Optogenetics Applications. *Photonics* **2021**, *8*, 220.
47. Steinmetz, N.; Koch, C.; Harris, K.; Carandini, M. Challenges and opportunities for large-scale electrophysiology with Neuropixels probes. *Curr. Opin. Neurobiol.* **2018**, *50*, 92–100.
48. Zhu, C.; Preissl, S.; Ren, B. Single-cell multimodal omics: The power of many. *Nat. Methods* **2020**, *17*, 11–14.
49. Tanay, A.; Regev, A. Scaling single-cell genomics from phenomenology to mechanism. *Nature* **2017**, *541*, 331–338.
50. Svensson, V.; Vento-Tormo, R.; Teichmann, S.A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **2018**, *13*, 599–604.
51. Regev, A.; Teichmann, S.A.; Lander, E.S.; Amit, I.; Benoist, C.; Birney, E.; Bodenmiller, B.; Campbell, P.; Carninci, P.; Clatworthy, M.; et al. The Human Cell Atlas. *eLife* **2017**, *6*, e27041.
52. Svensson, V.; de Veiga Beltrame, E.; Pachter, L. A curated database reveals trends in single-cell transcriptomics. *Database* **2020**, *2020*, baaa073.
53. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A.; et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **2009**, *6*, 377–382.
54. Chen, X.; Teichmann, S.A.; Meyer, K.B. From Tissues to Cell Types and Back: Single-Cell Gene Expression Analysis of Tissue Architecture. *Annu. Rev. Biomed. Data Sci.* **2018**, *1*, 29–51.
55. Ding, J.; Adiconis, X.; Simmons, S.K.; Kowalczyk, M.S.; Hession, C.C.; Marjanovic, N.D.; Hughes, T.K.; Wadsworth, M.H.; Burks, T.; Nguyen, L.T.; et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **2020**, *38*, 737–746.
56. Mereu, E.; Lafzi, A.; Moutinho, C.; Ziegenhain, C.; McCarthy Davis, J.; Álvarez-Varela, A.; Batlle, E.; Sagar, Grün, D.; Lau, J.K.; Boutet, S.C.; et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* **2020**, *38*, 747–755.
57. Zappia, L.; Phipson, B.; Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* **2018**, *14*, e1006245.
58. Hie, B.; Peters, J.; Nyquist, S.K.; Shalek, A.K.; Berger, B.; Bryson, B.D. Computational methods for single-cell RNA sequencing. *Annu. Rev. Biomed. Data Sci.* **2020**, *3*, 339–364.
59. Van den Berge, K.; Hembach, K.M.; Soneson, C.; Tiberi, S.; Clement, L.; Love, M.I.; Patro, R.; Robinson, M.D. RNA sequencing data: Hitchhiker’s guide to expression analysis. *Annu. Rev. Biomed. Data Sci.* **2019**, *2*, 139–173.
60. Ilany, A.; Holekamp, K.E.; Akçay, E. Rank-dependent social inheritance determines social network structure in spotted hyenas. *Science* **2021**, *373*, 348–352.
61. Bendall, S.C.; Davis, K.L.; Amir, E.D.; Tadmor, M.D.; Simonds, E.F.; Chen, T.J.; Shenfeld, D.K.; Nolan Garry, P.; Peer, D. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **2014**, *157*, 714–725.
62. La Manno, G.; Soldatov, R.; Zeisel, A.; Braun, E.; Hochgerner, H.; Petukhov, V.; Lidschreiber, K.; Kastrioti, M.E.; Lönnerberg, P.; Furlan, A.; et al. RNA velocity of single cells. *Nature* **2018**, *560*, 494–498.

63. Svensson, V.; Pachter, L. RNA Velocity: Molecular Kinetics from Single-Cell RNA-Seq. *Mol. Cell* **2018**, *72*, 7–9.
64. Loh, P.-R.; Baym, M.; Berger, B. Compressive genomics. *Nat. Biotechnol.* **2012**, *30*, 627–630.
65. Yu, Y.W.; Daniels, N.M.; Danko, D.C.; Berger, B. Entropy-Scaling Search of Massive Biological Data. *Cell Syst.* **2015**, *1*, 130–140.
66. Cleary, B.; Cong, L.; Cheung, A.; Lander, E.S.; Regev, A. Efficient Generation of Transcriptomic Profiles by Random Composite Measurements. *Cell* **2017**, *171*, 1424–1436.e18.
67. Stein-O'Brien, G.L.; Arora, R.; Culhane, A.C.; Favorov, A.V.; Garmire, L.X.; Greene, C.S.; Goff, L.A.; Li, Y.; Ngom, A.; Ochs, M.F.; et al. Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet.* **2018**, *34*, 790–805.
68. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
69. Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE* **2014**, *9*, e98679.
70. Weinreb, C.; Wolock, S.; Klein, A.M. SPRING: A kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **2018**, *34*, 1246–1248.
71. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2020**, arXiv:1802.03426.
72. Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.; Kwok, I.W.H.; Ng, L.G.; Ginhoux, F.; Newell, E.W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **2018**, *37*, 38–44. <https://doi.org/10.1038/nbt.4314>.
73. Kotliar, D.; Veres, A.; Nagy, M.A.; Tabrizi, S.; Hodis, E.; Melton, D.A.; Sabeti, P.C. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **2019**, *8*, e43803.
74. Nelson, W.; Zitnik, M.; Wang, B.; Leskovec, J.; Goldenberg, A.; Sharan, R. To Embed or Not: Network Embedding as a Paradigm in Computational Biology. *Front. Genet.* **2019**, *10*, 381.
75. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868.
76. Shafer, M.E.R. Cross-Species Analysis of Single-Cell Transcriptomic Data. *Front. Cell Dev. Biol.* **2019**, *7*, 175.
77. Chauvel, C.; Novoloaca, A.; Veyre, P.; Reynier, F.; Becker, J. Evaluation of integrative clustering methods for the analysis of multi-omics data. *Brief. Bioinform.* **2020**, *21*, 541–552.
78. Kiselev, V.Y.; Andrews, T.S.; Hemberg, M. Publisher Correction: Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **2019**, *20*, 310.
79. Zeng, W.; Chen, X.; Duren, Z.; Wang, Y.; Jiang, R.; Wong, W.H. DC3 is a method for deconvolution and coupled clustering from bulk and single-cell genomics data. *Nat. Commun.* **2019**, *10*, 4613.
80. Petegrosso, R.; Li, Z.; Kuang, R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. *Brief. Bioinform.* **2019**, *21*, 1209–1223.
81. Voelkl, B.; Altman, N.S.; Forsman, A.; Forstmeier, W.; Gurevitch, J.; Jaric, I.; Karp, N.A.; Kas, M.J.; Schielzeth, H.; van de Castele, T.; et al. Reproducibility of animal research in light of biological variation. *Nat. Rev. Neurosci.* **2020**, *21*, 384–393.
82. O'Leary, T.; Sutton, A.C.; Marder, E. Computational models in the age of large datasets. *Curr. Opin. Neurobiol.* **2015**, *32*, 87–94.
83. Cohen, D. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **1966**, *12*, 119–129.
84. Miguel, M.C.; Parley, J.T.; Pastor-Satorras, R. Effects of Heterogeneous Social Interactions on Flocking Dynamics. *Phys. Rev. Lett.* **2018**, *120*, 068303.
85. Knebel, D.; Ayali, A.; Guershon, M.; Ariel, G. Intra- versus intergroup variance in collective behavior. *Sci. Adv.* **2019**, *5*, eaav0695.
86. Stern, S.; Kirst, C.; Bargmann, C.I. Neuromodulatory Control of Long-Term Behavioral Patterns and Individuality across Development. *Cell* **2017**, *171*, 1649–1662.e10.
87. Stevenson-Hinde, J.; Zunz, M. Subjective assessment of individual rhesus monkeys. *Primates* **1978**, *19*, 473–48.
88. Mather, J.A.; Anderson, R.C. Personalities of octopuses (*Octopus rubescens*). *J. Comp. Psychol.* **1993**, *107*, 336–340.
89. Boring, L.; Gosling, J.; Cleary, M.; Charo, I.F. Decreased lesion formation in CCR2^{-/-} mice reveals a role for chemokines in the initiation of atherosclerosis. *Nature* **1998**, *394*, 894–897.
90. Forkosh, O. Animal behavior and animal personality from a non-human perspective: Getting help from the machine. *Patterns* **2021**, *2*, 100194.
91. Wice, E.W.; Saltz, J.B. Selection on heritable social network positions is context-dependent in *Drosophila melanogaster*. *Nat. Commun.* **2021**, *12*, 3357.
92. Bruijning, M.; Metcalf, C.J.E.; Jongejans, E.; Ayroles, J.F. The Evolution of Variance Control. *Trends Ecol. Evol.* **2020**, *35*, 22–33.
93. Hill, M.S.; Zande, P.V.; Wittkopp, P.J. Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **2021**, *22*, 203–215.
94. Dueck, H.; Eberwine, J.; Kim, J. Variation is function: Are single cell differences functionally important?: Testing the hypothesis that single cell variation is required for aggregate function. *BioEssays* **2016**, *38*, 172–180.
95. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **2015**, *25*, 1491–1498.
96. Arnol, D.; Schapiro, D.; Bodenmiller, B.; Saez-Rodriguez, J.; Stegle, O. Modeling Cell-Cell Interactions from Spatial Molecular Data with Spatial Variance Component Analysis. *Cell Rep.* **2019**, *29*, 202–211.e6.
97. Gustafsson, J.; Held, F.; Robinson, J.L.; Björnson, E.; Jörnsten, R.; Nielsen, J. Sources of variation in cell-type RNA-Seq profiles. *PLoS ONE* **2020**, *15*, e0239495.
98. Foreman, R.; Wollman, R. Mammalian gene expression variability is explained by underlying cell state. *Mol. Syst. Biol.* **2020**, *16*, e9146.

-
99. Osorio, D.; Yu, X.; Zhong, Y.; Li, G.; Yu, P.; Serpedin, E.; Huang, J.Z.; Cai, J.J. Single-Cell Expression Variability Implies Cell Function. *Cells* **2019**, *9*, 14.
 100. Shaffer, S.M.; Emert, B.L.; Reyes Hueros, R.A.; Cote, C.; Harmange, G.; Schaff, D.L.; Sizemore, A.E.; Gupte, R.; Torre, E.; Singh, A.; et al. Memory Sequencing Reveals Heritable Single-Cell Gene Expression Programs Associated with Distinct Cellular Behaviors. *Cell* **2020**, *182*, 947–959.e17.
 101. Phillips, N.E.; Mandic, A.; Omid, S.; Naef, F.; Suter, D.M. Memory and relatedness of transcriptional activity in mammalian cell lineages. *Nat. Commun.* **2019**, *10*, 1208.
 102. Javer, A.; Currie, M.; Lee, C.W.; Hokanson, J.; Li, K.; Martineau, C.N.; Yemini, E.; Grundy, L.J.; Li, C.; Ch'ng, Q.; et al. An open-source platform for analyzing and sharing worm-behavior data. *Nat. Methods* **2018**, *15*, 645–646.
 103. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **2018**, *36*, 411–420.
 104. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck William, M., 3rd; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e21.
 105. Wolf, F.A.; Angerer, P.; Theis, F.J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **2018**, *19*, 15.
 106. Branson, K.; Robie, A.; Bender, A.J.; Perona, P.; Dickinson, M.H. High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* **2009**, *6*, 451–457.
 107. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *21*, 1281–1289.
 108. Jezovit Jacob, A.; Alwash Nawar Levine Joel, D. Using Flies to Understand Social Networks. In *Frontiers in Neural Circuits*. **2021**, *15*, 1662–5110.
 109. Finn, K.R.; Silk, M.J.; Porter, M.A.; Pinter-Wollman, N. The use of multilayer network analysis in animal behaviour. *Anim. Behav.* **2019**, *149*, 7–22.
 110. Rocha, L.E.C.; Ryckebusch, J.; Schoors, K.; Smith, M. The scaling of social interactions across animal species. *Sci. Rep.* **2021**, *11*, 12584.
 111. Castles, M.; Heinsohn, R.; Marshall, H.H.; Lee Alexander, E.G.; Cowlshaw, G.; Carter Alecia, J. Social networks created with different techniques are not comparable. *Anim. Behav.* **2014**, *96*, 59–67.