Received: 4 May 2021 | Revised: 18 September 2021 | Accepted: 2 November 2021

DOI: 10.1002/bies.202100118

BioEssays

PROBLEMS & PARADIGMS

Prospects & Overviews

# Cell population-based framework of genetic epidemiology in the single-cell omics era

**Daigo Okada** | **Cheng Zheng** | **Jian Hao Cheng** | **Ryo Yamada**

Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto, Japan

**Correspondence**
Daigo Okada, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University,Nanbusogo-Kenkyu-To-1, 5F, 53 Syogoin-Kawaramachi, Sakyo-ku, Kyoto 606–8507, Japan.
Email: dokada@genome.med.kyoto-u.ac.jp

**Funding information**
Japan Science and Technology Agency, Grant/Award Numbers: AIP Challenge 2020, JPMJCR21U2; Japan Society for the Promotion of Science, Grant/Award Numbers: JP19J14816, 21K21316; Core Research for Evolutional Science and Technology, Grant/Award Numbers: JPMJCR1502, JPMJCR15G1

## Abstract

Genetic epidemiology is a rapidly advancing field due to the recent availability of large amounts of omics data. In recent years, it has become possible to obtain omics information at the single-cell level, so genetic epidemiological models need to be updated to integrate with single-cell expression data. In this perspective paper, we propose a cell population-based framework for genetic epidemiology in the single-cell era. In this framework, genetic diversity influences phenotypic diversity through the diversity of cell population profiles, which are defined as high-dimensional probability distributions of the state spaces of biomolecules of each omics layer. We discuss how biomolecular experimental measurement data can capture the different properties of this distribution. In particular, single-cell data constitute a sample from this population distribution where only some coordinate values are observable. From a data analysis standpoint, we introduce methodology for feature extraction from cell population profiles. Finally, we discuss how this framework can be applied not only to genetic epidemiology but also to systems biology.

**KEYWORDS**
genomics, genetics, single cell, transcriptome, epigenome, systems biology, GWAS

## INTRODUCTION

Understanding the phenotypic diversity among human populations is important in medicine and other life sciences. Genetic epidemiology evaluates phenotypic diversity by statistical models that combine genetic effects and environmental effects to identify the causal variants or genes of diseases. This has greatly contributed to the understanding of the genetic causes and mechanisms of disease.

In recent years, the field of genetic epidemiology has grown significantly due to the availability of genomics data. In particular, genome-wide association studies (GWAS) have identified many genetic variants that affect complex traits including diseases.[1] In addition to genomic information, information from other omics such as transcriptomics can also be used to analyze phenotypic diversity. Furthermore, in the past few years, the technology for measuring omics data at the single-cell level has made dramatic progress. The integration of genetic epidemiology methodology with single-cell omics data is becoming increasingly important. In this paper, we propose cell population-based frameworks and discuss the future of genetic epidemiology with single-cell omics data.

## MODEL FOR EXPLAINING THE VARIATION OF PHENOTYPE

### Standard model

The model that expresses phenotypic diversity as a combination of genetic and environmental effects is the most basic model in genetic

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 Wiley Periodicals LLC



15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

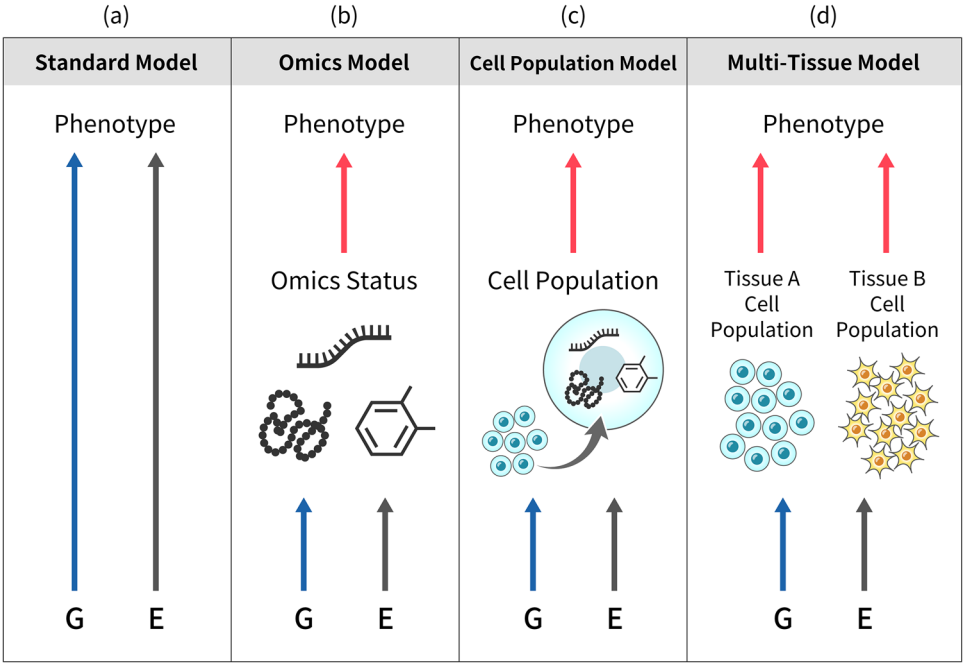


**FIGURE 1** Model of genetic epidemiology. G and E represent Genetic and Environmental effects, respectively. (A) Standard Model. G and E directly generate phenotypic diversity. (B) Omics Model. G and E generate phenotypic diversity via the diversity of omics information. (C) Cell Population Model. G and E generate phenotypic diversity via the diversity of cell populations where each single cell has omics information. (D) Multi-Tissue Model. G and E generate phenotypic diversity via the diversity of multiple cell populations

epidemiology (Standard Model: Figure 1A). Many genetic epidemiological studies, including GWAS, are based on this framework and use statistical models such as linear regression and contingency table tests to analyze the association of genetic factors and phenotype. This basic model expresses only the causal relationship from genetic factors to phenotypic diversity and does not include insight into molecular mechanisms.

## Omics model

Genetic factors influence phenotypic diversity of biomolecules such as RNA or proteins. Comprehensive biomolecular information is known as omics information, which is classified into genome, transcriptome, proteome, epigenome, or metabolome information.[2,3] Genetic epidemiologists have actively studied phenotypic diversity through such omics information, which is not limited to genomic information.

The Omics Model shown in Figure 1B is a framework for combining genetic epidemiology with omics data. In this model, genetic and environmental effects contribute to phenotypic diversity via biomolecular information. To identify the genetic effects on pools of biomolecular information such as the transcriptome, proteome, metabolome, or epigenome (blue arrow in Figure 1B), the identification of single nucleotide polymorphisms (SNPs) associated with these biomolecules (expression quantitative trait loci (eQTL), protein QTL, methylation QTL, metabolite QTL) are being actively investigated.[4,5] For example, eQTL analysis identifies genetic variants that are associated with gene expression levels obtained from transcriptome data. The eQTLs identified in various tissues have been published in databases such as.[6,7] In addition, studies that examine the relationship between omics diversity and phenotypic diversity (red arrow in Figure 1B) constitute disease omics analysis. Studies to identify differentially expressed genes in diseased and healthy individuals are included in this category. Both types of study designs have been widely implemented in omics research projects.

## Cell population model

A disease or complex phenotype of medical interest is manifest at the tissue or individual level. It is not caused by just one particular cell but by abnormalities of an entire cell population in the relevant tissue. In fact, tissue samples used in omics analyses are composed of a number of cells, and each cell has different omics information. Breakthroughs in single-cell omics technology over the last few years have made it possible to acquire omics information at the single-cell level. Genetic epidemiological models can then be extended for single-cell omics studies.

We propose the Cell Population Model as a framework for genetic epidemiology with single-cell omics data (Figure 1C). This model expresses phenotypic diversity as cell population profile diversity. Each cell in the body has a different omics status from the others. In this model, genetic and environmental effects affect phenotypic diversity through the diversity of a cell population profile where each cell has omics information. This model is an extension of the Omics Model and is considered a natural biological expression of complex phenotypes.

15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

While association studies between cell population profiles and phenotype are often performed to identify a cellular subset related with disease using cytometry data or single-cell RNA-seq data (red arrow in Figure 1C),[8–10] genetic epidemiological analyses based on such models have not been performed to date (blue arrow in Figure 1C). Previously, we performed the first GWAS study on the diversity of lymphocyte populations in peripheral blood using a large-scale cytometry dataset based on this framework.[11] As a result, although the analysis was performed with a relatively small sample size, the SNPs associated with individual differences of the lymphocyte profile were successfully identified. In recent years, research to acquire cytometry data on a large scale has also become common.[12] Genetic epidemiological research under this model can be expected to bring new findings.

### Multi-tissue model

The Cell Population Model can be extended to multiple tissues in the Multi-Tissue Model (Figure 1D). Under this framework, phenotypic diversity is understood as being generated by a combination of effects from cell population profiles of multiple related tissues. For systemic diseases involving multiple tissues, such models are a natural expression of the mechanism. Although genetic epidemiological studies using the Multi-Tissue Model have not been conducted, it is considered meaningful as a future genetic epidemiological model in the single-cell era.

## CELL POPULATION PROFILE AS A DISTRIBUTION ON OMICS STATE SPACE

### Omics state space

Each cell in a cell population has the biomolecular information of five omics layers: epigenome, transcriptome, proteome, metabolome, and somatic genome. Biomolecular information in the epigenome layer, such as DNA methylation, histone acetylation, and chromatin openness, can be quantified as signal values assigned to each position in the genome. The factors in the transcriptome layer are the expression level of all genes in the human genome. The factors in the proteome layer are the expression levels of all proteins. More dimensions are required if cellular localization or chemical modifications such as phosphorylation of proteins are distinguished. The metabolome layer contains the abundance of all metabolites including lipids and low-molecular-weight compounds. Factors in the somatic genome layer are information about mutations or DNA damage that accumulate in the somatic genome and are distinct from the germline genome information inherited from parents. For example, cancer is a disease caused by an increase in the number of cells with abnormal somatic genomic information, and cancer genome analysis has been used to identify genes involved in the pathogenesis of the disease.[13,14] In addition, considering mitochondria genome is beneficial to understand the differences among cells. For example, recent in vivo study using mouse observed the mitochondrial transfer between different types of cells, which is related to biological or pathological phenomena.[15,16]

Because each cell has individual omics information, one cell can be represented as a one point in the state space where each biomolecule measurement value represents a coordinate axis. Here, we call this state space of the biomolecules of all the omics layers the "Omics State Space." The function of the cell population depends on the profile of cells with different omics statuses.

Therefore, the cell population profile is characterized as a probability distribution in the Omics State Space. Since this distribution corresponds to the joint distribution of whole biomolecular measurements, including all gene expressions, protein expressions, mutations in the somatic genome, and epigenome modifications, it is a very high-dimensional distribution. Cells are not evenly observed in the Omics State Space, and most parts are sparse areas where no cells are observed at all. We define the cell population profile as the distribution in this Omics State Space.

## EXPERIMENTAL DATA MEASURING BIOMOLECULES TO CAPTURE THE PROPERTIES OF THE CELL POPULATION PROFILE

Experimental data measuring biomolecules can be interpreted as capturing different parts of the distribution of the cell population profile in the Omics State Space. Because the distribution of cell population profiles is very high-dimensional and complex, there is no experiment technique to get a complete picture. Existing biomolecular experimental data can be classified according to three perspectives with respect to the desired information: the target omics layer, bulk/single-cell, and candidate-based/comprehensive. For example, bulk and candidate-based approaches in the proteome layer include western blotting or ELISA. Immunocytochemistry is a single-cell level and candidate-based method primarily in the proteome layer, where the number of cells that can be measured is small but protein localization can be distinguished. Single-cell and comprehensive approaches in the transcriptome layer include RNA-seq or DNA microarray. Methods for comprehensive measurements at the single-cell level in each layer have made rapid progress in the past few years.[15–22] Recent genomics assay can detect even mtDNA mutations at single cell level.[23]

In particular, single-cell data and bulk data differ in their data structure. The bulk measurement is an estimate of the mean value for a particular axis of the distribution in the Omics State Space. Since the mean value in a probability distribution is a representative and reasonable feature of the distribution, the bulk measurement value is a reasonable index for comparison among distributions when the cell populations are homogeneous. Single-cell data constitute a sample from the population distribution of the cell population profile where only some coordinate values are observable (Figure 2). While the shape information of the distribution is lost in bulk data, single-cell data can partially capture it. Then, it can be used to identify and quantify heterogeneity and cellular subsets in the cell population profile.

15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

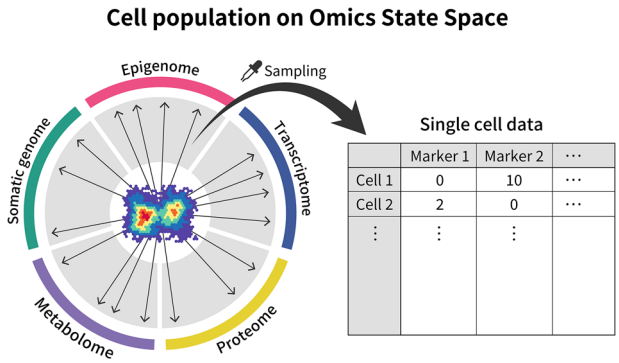


**FIGURE 2** Omics State Space and single cell data. The cell population profile is characterized as a high-dimensional probability distribution in the Omics State Space where each measurement value on transcriptome, proteome, metabolome, somatic genome, or epigenome layer represents a coordinate axis. Single-cell data constitute a sample from this population distribution of the cell population profile where only some coordinate values are observable as marker

The ability to acquire more biomolecule information simultaneously at the single-cell level will allow us to understand the shape of the cell population profile at higher resolution. In recent years, the ability to measure omics information in multiple layers simultaneously has been actively researched, and measurement techniques at the single-cell level have been developed.[24,25]

## REQUIREMENT OF A CELL POPULATION

In this section, we will discuss important issues when considering cell population profiles as distributions, and the requirements that must be met for a cell population.

When the cell population profile is viewed as a probability distribution, each data point is considered independent and the cell location information disappears. Then, cells need to be able to come and go from each other within a cell population. This assumption holds well for peripheral blood cell populations. When blood cells are sampled from peripheral blood, each cell can be regarded as independent and randomly collected, and the single-cell data can be regarded as a statistical sample from the population distribution. However, in many anatomically defined tissues, it is not only necessary for cells to maintain their proper biomolecular expression state, but also for each cell to occupy its proper position in the tissue to maintain the tissue function. For example, tissue stem cells are maintained in a microenvironment called a niche.[26] Considering such cell populations as distributions would result in a loss of biological information.

In recent years, spatial omics technologies that simultaneously acquire positional and omics information have received much attention. For example, the spatial transcriptome can reveal transcriptome data while retaining spatial information in the tissue.[27] Such spatial information may be useful in determining the range of cell populations that can be treated as distributions and in compensating for the loss of positional information.

To extend the Cell Population Model to the Multi-Tissue Model, it is necessary to consider the interactions between the cell populations. Cell populations exchange information through physical interactions or cellular signaling. In reality, the diversity of some complex phenotypes is generated by many cell populations that make up an individual and their interactions.

## FEATURE EXTRACTION OF CELL POPULATION PROFILES

In order to design genetic epidemiology studies based on a cell population-based framework, such as the Cell Population Model or Multi-Tissue Model, it is necessary to perform association analysis between the cell population profile and individual labels such as genotype or phenotype. Since the cell population profile is represented as a probability distribution on the Omics State Space, conventional methods of genetic epidemiology and omics data analysis cannot be directly used in this situation. The solution to apply these data analysis methods and conduct association analysis is to extract feature values from cell population profiles. In this section, we introduce three conventional ideas on feature extraction of cell population profiles, methods using bulk data, methods based on cellular subsets, and non-parametric methods.

The mean value of distribution obtained by bulk data is one of the most commonly used features of cell population profiles. For example, bulk transcriptome data have contributed greatly to the identification of tissue-specific genes [28]. The identification of tissue-specific genes is way to compare cell populations from multiple tissues to find transcriptome axes whose mean values differ significantly among the multiple tissue cell populations on the Omics State Space. In the medical science field, many searches for biomolecular markers using bulk data have been conducted.[29,30]

Feature extraction based on cellular subsets is frequently done with single-cell data. Each cell in a cell population is a little different from the others, so no two cells are exactly the same. However, since cell populations are formed as cells proliferate and differentiate, there are cellular subsets with the same properties and functions in the cell population. Therefore, we can understand cellular function by classifying cells into subsets and annotating their functions. Since a cell population profile is a mixed distribution of cellular subsets, a quantitative value of the percentage of each subset is also a valid feature of a cell population profile. Computational methods for clustering cells using single-cell data to identify cellular subsets are actively being studied by computational biologists.[31,32]

Cellular subset-based feature extraction also loses information. One reason is that the results of feature extraction are affected by prior biological knowledge and assumptions about the pre-identified cellular subsets. However, it is not known exactly how many cellular subsets there are in our body or how we should classify them. Novel subsets are being newly identified. Even data-driven classification using information science methods cannot eliminate such biases due to the assumptions made in the algorithms and statistical models. In addition,

15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

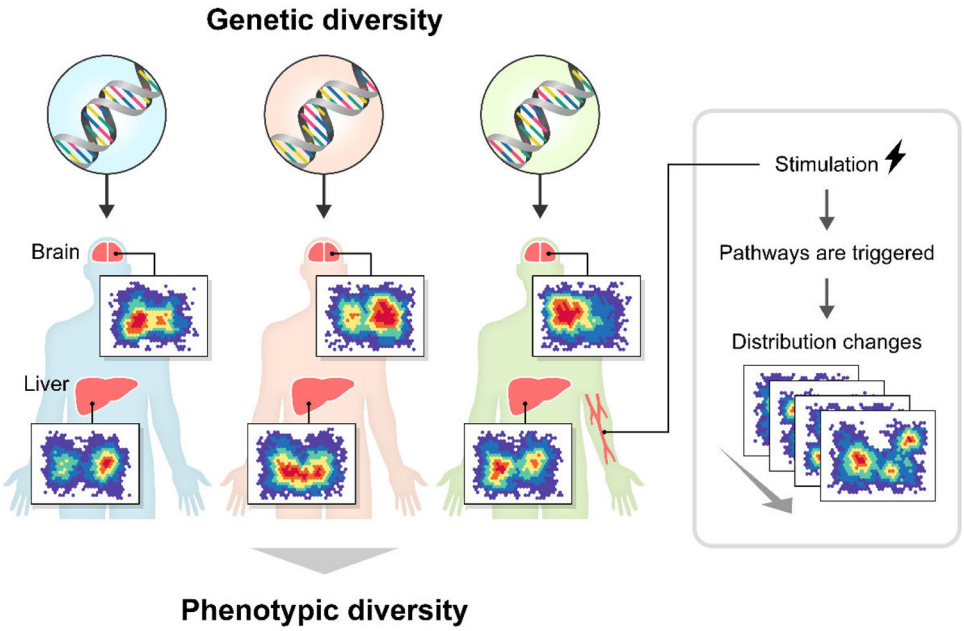


**FIGURE 3** Graphical abstract of cell population-based framework integrating genetic epidemiology and systems biology. In the context of genetic epidemiology, genetic effects influence phenotypic diversity through their impact on the probability distributions of various cell population profiles that make up an individual. In the context of systems biology, within an individual, responses to stimuli drive biological pathways and cause biological phenomena by altering their distribution. In both cases, extracting feature values from the distribution makes it possible to represent them in a statistical model

information about variability and diversity within cellular subsets is also lost.

Nonparametric feature extraction is another means to obtain feature values without assumptions about cellular subsets. A nonparametric statistical method models the probability distribution of the cell population profile without caring about the number of parameters. In cytometry data analysis, a method using information theory-based dissimilarity quantification and multi-dimensional scaling (MDS) has been proposed.[33,34] Here, the dissimilarity matrix among probability distributions is calculated by nonparametric density estimation, and MDS is applied to this dissimilarity matrix to obtain coordinates that reflect the dissimilarity relationship. Decomposition into Extended Exponential Family expresses the cell population profile distribution as an exponential family-like formula in a nonparametric manner, giving coordinates based on the inner-product matrix among the distributions.[35] The coordinates obtained by these procedures can be treated as data-driven feature values of the cell population profiles.

The development of feature extraction methods that satisfy these requirements is a future challenge in data analysis for implementing genetic epidemiology models in the single-cell era. The advantage of cellular subset-based feature extraction is that the biological meaning of the obtained features is clear and easy to interpret. The advantage of nonparametric methods is that they can model cell population profiles without using prior assumptions about cellular subsets. However, nonparametric methods generally require larger sample sizes to perform robust analysis. Due to cost issues, it is often difficult to acquire single-cell data with very large sample sizes. While there are many methods to compare multiple samples in cytometry data, such methods are lacking in single-cell RNA-seq data in particular.[36] That is a future task in single cell data analysis.

## DYNAMICS OF CELL POPULATION PROFILES

Research designs that combine genetic epidemiology and systems biology are useful in medical research.[37,38] In genetic epidemiology, genetic factors are identified by analyzing the diversity of complex phenotypes such as diseases among individuals. However, understanding the dynamics of complex phenotypes within the same individual is also important for the control of diseases with a systems biology approach. For example, how the body responds to drug stimuli and their molecular mechanisms is a fundamental research topic in the medical sciences. The dynamics of biological phenomena can also be explained by a cell population-based framework.

The cell population profile changes over time in response to external stimuli to affect biological phenomena. For example, lymphocyte populations in the peripheral blood change after vaccination to cause an immune response. In addition, external signals trigger cellular responses such as proliferation and differentiation. Therefore, the state of a cell at a future time point depends on the current Omics state and the external environment of the cell. This transition rule is defined by the biological pathway.

Because of each cell's dynamics, the overall distribution changes when a cell at a certain coordinate moves to another point depending on the type of stimulus. Cell death, division, or proliferation can also cause the change in distribution. This change is triggered by a change

15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in the proteome layer, in which receptor proteins on the cell surface respond to an external substance and change their activity. The regulatory relationships between biomolecules determine where cells move from one point to another in the Omics State Space depending on the presence and type of stimulation.

The dynamics of such cell population profiles can also be analyzed by applying ordinary data analysis methods through feature extraction. The changes in the cell population profiles for each sample can be visualized and analyzed as the time series data of its features. Various data analysis methods are available to analyze time series data.[39] A cell population-based framework such as the Cell Population Model or Multi-Tissue Model provides an integrated approach to both genetic epidemiology and systems biology (Figure 3). This framework will be useful to investigate the genetic effect that is condition-specific or related with dynamics. For example, the recent research suggested that the RNASEH2B variant has relation to hemophagocytic lymphohistiocytosis (HLH) depending on the biological condition.[40]

## CONCLUSION

In this perspective paper, we proposed a cell population-based framework for genetic epidemiology. In this framework, genetic diversity influences phenotypic diversity through the diversity of cell population profiles. Cell population profiles are high-dimensional distributions on the Omics State Space, and all biomolecular measurement data are used to obtain the properties of this distribution. To conduct genetic epidemiology in a cell population-based framework, feature extraction from cell population profiles is important from a data analysis standpoint. In addition, this framework can also be applied to represent the dynamics of cell population profiles, providing an integrated approach to genetic epidemiology and systems biology.

### ACKNOWLEDGMENTS

This work was supported by a KAKENHI Grant-in-Aid from the Japan Society for the Promotion of Science (JSPS; grant number JP19J14816 and 21K21316), the Core Research for Evolutionary Science and Technology (CREST; grant numbers JPMJCR1502 and JPMJCR15G1), and the Japan Science and Technology Agency (JST; the AIP Challenge and JPMJCR21U2).

### CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

### ORCID

*Daigo Okada* https://orcid.org/0000-0002-9725-6373

### REFERENCES

1. Marees, A. T., De Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, *27*(2), e1608.
2. Vilne, B., & Schunkert, H. (2018). Integrating genes affecting coronary artery disease in functional networks by multi-omics approach. *Frontiers in Cardiovascular Medicine*, *5*, 89
3. Yan, J., Risacher, S. L., Shen, L., & Saykin, A. J. (2018). Network approaches to systems biology analysis of complex disease: Integrative methods for multiomics data. *Briefings in Bioinformatics*, *19*(6), 1370–1381.
4. Wen, J., Nodzak, C., & Shi, X. (2020). Qtl analysis beyond eQTLs. *Methods in Molecular Biology*, *2082*, 201–210.
5. Schlosser, P., Li, Y., Sekula, P., Raffler, J., Grundner-Culemann, F., Pietzner, M., Cheng, Y., Wuttke, M., Steinbrenner, I., Schultheiss, U. T., Kotsis, F., Kacprowski, T., Forer, L., Hausknecht, B., Ekici, A. B., Nauck, M., Völker, U., Walz, G., Oefner, P. J., & Köttgen, A. (2020). Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nature Genetics*, *52*(2), 167–176.
6. Narahara, M., Higasa, K., Nakamura, S., Tabara, Y., Kawaguchi, T., Ishii, M., Matsubara, K., Matsuda, F., & Yamada, R. (2014). Large-scale east-asian eqtl mapping reveals novel candidate genes for ld mapping and the genomic landscape of transcriptional effects of sequence variants. *PLoS ONE*, *9*(6), e100924. https://doi.org/10.1371/journal.pone.0100924.
7. GTEx Consortium et al. (2017). Genetic effects on gene expression across human tissues. *Nature*, *550*(7675), 204–213.
8. Arvaniti, E., & Claassen, M. (2017). Sensitive detection of rare disease-associated cell subsets via representation learning. *Nature Communications*, *8*(1), 1–10.
9. Bailur, J. K., Mccachren, S. S., Doxie, D. B., Shrestha, M., Pendleton, K., Nooka, A. K., Neparidze, N., Parker, T. L., Bar, N., Kaufman, J. L., Hofmeister, C. C., Boise, L. H., Lonial, S., Kemp, M. L., Dhodapkar, K. M., & Dhodapkar, M. V. (2019). Early alterations in stem-like/marrow-resident *t* cells and innate and myeloid cells in preneoplastic gammopathy. *JCI Insight*, *4*, 11.
10. Boland, B. S., He, Z., Tsai, M. S., Olvera, J. G., Omilusik, K. D., Duong, H. G., Kim, E. S., Limary, A. E., Jin, W., Milner, J. J., Yu, B., Patel, S. A., Louis, T. L., Tysl, T., Kurd, N. S., Bortnick, A., Quezada, L. K., Kanbar, J. N., & Chang, J. T. (2020). Heterogeneity and clonal relationships of adaptive immune cells in ulcerative colitis revealed by single-cell analyses. *Science Immunology*, *5*, 50.
11. Okada, D., Nakamura, N., Setoh, K., Kawaguchi, T., Higasa, K., Tabara, Y., Matsuda, F., & Yamada, R. (2021). Genome-wide association study of individual differences of human lymphocyte profiles using large-scale cytometry data. *Journal of Human Genetics*, *66*, 557–567.
12. Tsang, J. S., Schwartzberg, P. L., Kotliarov, Y., Biancotto, A., Xie, Z., Germain, R. N., Wang, E., Olnes, M. J., Narayanan, M., Golding, H., Moir, S., Dickler, H. B., Perl, S., Cheung, F., Obermoser, G., Chaussabel, D., Palucka, K., Chen, J., Fuchs, J. C., Young, N. S. (2014). Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell*, *157*(2), 499–513.
13. Mallory, X. F., Edrisi, M., Navin, N., & Nakhleh, L. (2020). Methods for copy number aberration detection from single-cell dna-sequencing data. *Genome Biology*, *21*(1), pp. 1–22.
14. Barrett, I. P. (2010). Cancer genome analysis informatics. *Genetic Variation. Methods in Molecular Biology (Methods and Protocols)*, *68*, 75–102.
15. Brestoff, J. R., Wilen, C. B., Moley, J. R., Li, Y., Zou, W., Malvin, N. P., Rowen, M. N., Saunders, B. T., Ma, H., Mack, M. R., Hykes, B. L., Balce, D. R., Orvedahl, A., Williams, J. W., Rohatgi, N., Wang, X., Mcallaster, M. R., Handley, S. A., Kim, B. S., & Teitelbaum, S. L. (2021). Intercellular mitochondria transfer to macrophages regulates white adipose tissue homeostasis and is impaired in obesity. *Cell Metabolism*, *33*(2), 270–282.e8.
16. Moschoi, R., Imbert, V., Nebout, M., Chiche, J., Mary, D., Prebet, T., Saland, E., Castellano, R., Pouyet, L., Collette, Y., Vey, N., Chabannon, C., Recher, C., Sarry, J. -. E., Alcor, D., Peyron, J.-F., & Griessinger, E. (2016). Protective mitochondrial transfer from bone marrow stromal cells to

15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

acute myeloid leukemic cells during chemotherapy. *Blood*, *128*(2), 253–264.
17. Islam, M., Chen, B., Spraggins, J M., Kelly, R T., & Lau, K S. (2020). Use of single cell-omic technologies to study the gastrointestinal tract and diseases, from single cell identities to patient features. *Gastroenterology*, *159*(2), 453–466 e1.
18. Kumar, R., Ghosh, M., Kumar, S., & Prasad, M. (2020). Single cell metabolomics: A future tool to unmask cellular heterogeneity and virus-host interaction in context of emerging viral diseases. *Frontiers in Microbiology*, *11*(1152), https://doi.org/10.3389/fmicb.2020.01152.
19. Mimitou, E. P., Lareau, C. A., Chen, K. Y., Zorzetto-Fernandes, A. L., Hao, Y., Takeshima, Y., & Smibert, P. (2021). Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology*, *39*, 1246–1258.
20. Chen, H., Lareau, C., Andreani, T., Vinyard, M E., Garcia, S P., Clement, K., Andrade-Navarro, M A., Buenrostro, J D., & Pinello, L. (2019). Assessment of computational methods for the analysis of single-cell atac-seq data. *Genome Biology*, *20*(1), 1–25.
21. Rotem, A., Ram, O., Shoresh, N., Sperling, R. A., Goren, A., Weitz, D. A., & Bernstein, B. E. (2015). Single-cell chip-seq reveals cell subpopulations defined by chromatin state. *Nature Biotechnology*, *33*(11), 1165–1172.
22. Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyal, F., Frenoy, O., Pousse, Y., Reichen, M., Woolfe, A., Brenan, C., Griffiths, A. D., Vallot, C., & Gérard, A. (2019). High-throughput single-cell chip-seq identifies heterogeneity of chromatin states in breast cancer. *Nature Genetics*, *51*(6), 1060–1066.
23. Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., Pelka, K., Ge, W., Oren, Y., Brack, A., Law, T., Rodman, C., Chen, J. H., Boland, G M., Hacohen, N., Rozenblatt-Rosen, O., Aryee, M. J., Buenrostro, J. D., Regev, A., & Sankaran, V G. (2019). Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell*, *176*(6), 1325–1339.e22.
24. Hu, Y., An, Q., Sheu, K., Trejo, B., Fan, S., & Guo, Y. (2018). Single cell multiomics technology: Methodology and application. *Frontiers in Cell and Developmental Biology*, *6*(28), https://doi.org/10.3389/fcell.2018.00028.
25. Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*, *18*(1), 1–15.
26. Scadden, D. T. (2006). The stem-cell niche as an entity of action. *Nature*, *441*(7097), 1075–1079.
27. Moor, A. E., & Itzkovitz, S. (2017). Spatial transcriptomics: Paving the way for tissue-level systems biology. *Current Opinion in Biotechnology*, *46*, 126–133.
28. Liu, X., Yu, X., Zack, D. J., Zhu, H., & Qian, J. (2008). Tiger: A database for tissue-specific gene expression and regulation. *BMC Bioinformatics*, *9*(1), 1–7.
29. Han, J., Chen, M., Wang, Y., Gong, B., Zhuang, T., Liang, L., & Qiao, H. (2018). Identification of biomarkers based on differentially expressed genes in papillary thyroid carcinoma. *Scientific Reports*, *8*(1), 1–11.
30. Akond, Z., Alam, M., & Mollah, M. N. H (2018). Biomarker identification from rna-seq data using a robust statistical approach. *Bioinformation*, *14*(4), 153–163.
31. Saeys, Y., Van Gassen, S., & Lambrecht, B. N. (2016). Computational flow cytometry: Helping to make sense of high-dimensional immunology data. *Nature Reviews Immunology*, *16*(7), 449–462.
32. Kiselev, V Yu, Andrews, T S., & Hemberg, M. (2019). Challenges in unsupervised clustering of single-cell rna-seq data. *Nature Reviews Genetics*, *20*(5), 273–282.
33. Carter, K. M., Raich, R., Finn, W. G., & Hero, A. O. (2009). Fine: Fisher information nonparametric embedding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(11), 2093–2098.
34. Gingold, J. A., Coakley, E. S., Su, J., Lee, D. -. F., Lau, Z., Zhou, H., Felsenfeld, D. P., Schaniel, C., & Lemischka, I. R. (2015). Distribution analyzer, a methodology for identifying and clustering outlier conditions from singlecell distributions, and its application to a nanog reporter rnai screen. *BMC Bioinformatics*, *16*(225), https://doi.org/10.1186/s12859-015-0636-7.
35. Okada, D., & Yamada, R. (2020). Decomposition of a set of distributions in extended exponential family form for distinguishing multiple oligodimensional marker expression profiles of single-cell populations and visualizing their dynamics. *PLoS ONE*, *15*(4), e0231250, https://doi.org/10.1371/journal.pone.0231250.
36. Lähnemann, D., Köster, J., Szczurek, E., Mccarthy, D. J., Hicks, S. C., Robinson, M. D., Vallejos, C A., Campbell, K. R., Beerenwinkel, N., Mahfouz, A., Pinello, L., Skums, P., Stamatakis, A., Attolini, C. S.-.O., Aparicio, S., Baaijens, J., Balvert, M., Barbanson, B. D., Cappuccio, A., & Schönhuth, A. (2020). Eleven grand challenges in single-cell data science. *Genome Biology*, *21*(1), 1–35.
37. Li, H. (2013). Systems biology approaches to epidemiological studies of complex diseases. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *5*(6), 677–686.
38. Campbell, T. M., Castro, M. A. A., De Santiago, I., Fletcher, M. N. C., Halim, S., Prathalingam, R., Ponder, B. A. J., & Meyer, K. B. (2016). Fgfr2 risk snps confer breast cancer risk by augmenting oestrogen responsiveness. *Carcinogenesis*, *37*(8), 741–750.
39. Fu, T. C. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence*, *24*(1), 164–181.
40. Prokop, J. W., Shankar, R., Gupta, R., Leimanis, M. L., Nedveck, D., Uhl, K., Chen, B., Hartog, N. L., Van Veen, J., Sisco, J. S., Sirpilla, O., Lydic, T., Boville, B., Hernandez, A., Braunreiter, C., Kuk, C. C., Singh, V., Mills, J., Wegener, M., & Rajasekaran, S. (2020). Virus-induced genetics revealed by multidimensional precision medicine transcriptional workflow applicable to COVID-19. *Physiological Genomics*, *52*(6), 255–268.

**How to cite this article:** Okada, D., Zheng, C., Cheng, J. H., & Yamada, R. (2022). Cell population-based framework of genetic epidemiology in the single-cell omics era. *BioEssays*, *44*, e2100118. https://doi.org/10.1002/bies.202100118

15211878, 2022, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bies.202100118 by Cochrane Japan, Wiley Online Library on [08/12/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License