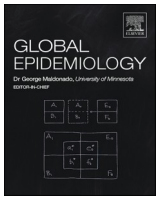
Global Epidemiology 8 (2024) 100156

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology



Conceptual frameworks for the integration of genetic and social epidemiology in complex diseases

Diane Xue [a,*], Anjum Hajat [b], Alison E. Fohner [a,b]

[a] Institute for Public Health Genetics, University of Washington School of Public Health, 1959 NE Pacific St, Room H-690, Seattle, WA 98195, USA
[b] Department of Epidemiology, University of Washington School of Public Health, Hans Rosling Population Health Building, 3980 15th Ave NE, Seattle, WA 98195, USA

A B S T R A C T

Uncovering the root causes of complex diseases requires complex approaches, yet many studies continue to isolate the effects of genetic and social determinants of disease. Epidemiologic efforts that under-utilize genetic epidemiology methods and findings may lead to incomplete understanding of disease. Meanwhile, genetic epidemiology studies are often conducted without consideration of social and environmental context, limiting the public health impact of genomic discoveries. This divide endures despite shared goals and increases in interdisciplinary data due to a lack of shared theoretical frameworks and differing language. Here, we demonstrate that bridging epidemiological divides does not require entirely new ways of thinking. Existing social epidemiology frameworks including Ecosocial theory and Fundamental Cause Theory, can both be extended to incorporate principles from genetic epidemiology. We show that genetic epidemiology can strengthen, rather than detract from, efforts to understand the impact of social determinants of health. In addition to presenting theoretical synergies, we offer practical examples of how genetics can improve the public health impact of epidemiology studies across the field. Ultimately, we aim to provide a guiding framework for trainees and established epidemiologists to think about diseases and complex systems and foster more fruitful collaboration between genetic and traditional epidemiological disciplines.

Introduction

Many of the leading causes of death have complex etiology. Cardiovascular disease, cancer, and Alzheimer's disease, for example, are influenced by heterogeneous genetic and environmental risk factors. Interdisciplinary tools and perspectives are needed to understand the underlying biology and develop effective prevention and treatment strategies for complex diseases, yet the integration of genetic and social epidemiology has remained limited.

Although there is a shared goal across epidemiology to uncover the causes of disease and prevent their occurrence, few researchers are adequately trained to understand the nuances of interdisciplinary approaches or translate between genetic and non-genetic subfields of epidemiology. There are several reasons for the continued siloing of genetic and non-genetic epidemiological subfields: both subfields train researchers to think first and foremost about isolating main effects rather than interactions; a history of biological essentialism in genetics has devalued the inclusions of other types of data; and historically, few datasets have captured genetic and social data for diverse populations.

However, data and methods have progressed. Large-scale cohort studies collecting multifaceted biomarkers and social/environmental data like *All of Us*, the *National Longitudinal Study of Adolescent to Adult Health (AddHealth),* and *Framingham Heart Study* include hundreds of variables across biological and social measures for ever-growing sample sizes [1–3]. Paired with the latest statistical and computational advancements, we have better capacity to handle high-dimensional, multilevel data analysis. Now is a pertinent time to revisit conceptual frameworks underlying epidemiological research and move toward a more interdisciplinary conceptualization of risk for complex diseases.

Theory and conceptual frameworks help epidemiologists evaluate the causes of health and disease, thereby influencing research questions, collaborations, study design and interpretation of findings [4,5]. While long-standing social epidemiological theories, such as Ecosocial theory and Fundamental Cause theory, describe hierarchical relationships between multi-level determinants of health, these frameworks do not deliberate on the role of genomics and other -omics on health outcomes [6,7]. Similarly, the field of genetic epidemiology has not fully embraced the use of conceptual models that could place genetics in the context of other determinants of health.

Here, we address the critical need for frameworks that highlight the parallels between genetic and social epidemiology and provide examples for potential collaboration. Rather than reinventing the wheel, we demonstrate that genetic epidemiology principles already play a hidden role in Ecosocial theory and Fundamental Cause theory. By bringing

* Corresponding author.
E-mail addresses: dxue@uw.edu (D. Xue), anjumh@uw.edu (A. Hajat), afohner@uw.edu (A.E. Fohner).

https://doi.org/10.1016/j.gloepi.2024.100156
Received 30 January 2024; Received in revised form 11 June 2024; Accepted 6 July 2024
Available online 8 July 2024

these genetic concepts to the forefront, we show that there is no need for contention between genetic and social explanations of health outcomes, and continuing to study them in isolation will only lead to incomplete understandings of disease. Genetic epidemiology tools can reinforce the importance of social determinants of health by improving precision in measuring effects of contextual determinants and refining tests of fundamental causes that underly persistent health disparities.

While our main priority is to map out the theoretical synergies between subfields that have been viewed in opposition in the past, we also offer a practical guide for translating theory into practice. We compiled a series of guiding questions to illustrate how to incorporate genomic information alongside social determinants into each stage of the epidemiological research process. In addition to the questions to guide interdisciplinary thinking, we have provided references to related literature to facilitate the application of these approaches. From study design to dissemination of results, including both genetic and social epidemiology principles and methods can improve public health impact and help genomics research move toward health equity (Table 1). The suggestions provided are non-exhaustive, and we hope this commentary provides a scaffold for active discussion going forward.

Previous integrated frameworks

Several conceptual frameworks have aimed to integrate social and biological determinants in understanding health disparities. Glass and McAtee's society-behavior-biology nexus emphasizes the dynamic interactions between social context and biological phenomena, with a focus on health behaviors [20]. While this model offers a useful theory suggesting potential synergy between underwater 'genetic substrate' and uphill social determinants, the theory does not take a deep dive — to borrow their metaphor – into the genetic epidemiology of disease, focusing more on the social conditions.

On the other hand, Boardman et al. and Shanahan and Hofer offer different typologies of GxE interactions [16,21], carefully disentangling different mechanisms and interpretations of interaction. These key contributors to complex systems literature discusses the potential of genetics for understanding health behavior and echoes theories of fundamental causes and embodiment to define the environment in GxE, but do not discuss how genetics and other -omics data fits into theoretical or conceptual understandings of disease.

Recent frameworks of social explanations of disease take advantage of -omics data, particularly the epigenome. One such example is Shantz and Elliot's focus on the 'social epigenome [22].' Like Glass and McAtee and Boardman et al., Shantz and Elliot also draw on Ecosocial theory and fundamental causes. They explain how the epigenome can mediate effects of place-based social variation, but do not discuss the role of genetic variation.

In 2012, Diez Roux described four conceptual approaches to study health disparities, focusing on genetic drivers, fundamental causes, pathways of embodiment, and gene-environment interactions [23]. While these approaches are initially explained separately, the paper calls for a complex systems framework that integrates gene-environment interplay with distal or fundamental causes and mediating pathways, noting that such frameworks may require new data and new ways of thinking. Diez Roux also cautions that linking genetic variation with health disparities, especially those concerning race and ethnicity, may inadvertently reify biological definitions of race. We strongly heed these words of caution, describing how genetic variation can be included in health disparities research without supporting a genetic basis for social stratification.

All of the aforementioned work makes important contributions to a more complete conceptual understanding of complex disease, but a more thorough integration of genetic epidemiology into theories of health disparities is still needed. It is no coincidence that Fundamental Cause Theory and pathways of embodiment are central to all of these previous models. They remain foundational in epidemiological research,

Table 1
Guidance for Integrating Genetic and Social Epidemiology Across Research Stages. This table presents a framework for interdisciplinary collaboration between genetic and social epidemiology throughout the research process. Each stage includes key questions to guide the integration of genomic information with social determinants, along with references to literature that provide additional insights and examples.

Research Stage	Key Considerations for Integrating Genetic and Social Epidemiology	Supporting Literature
Study Design	• What are the genetic and environmental risk factors associated with the trait? Is the trait influenced by a single gene (monogenic), several genes (polygenic), or an interplay of genetic and environmental risks? • Does genetic variation act as a confounder or an effect modifier of social or environmental exposures?	• The liability threshold model explains how genetic and environmental factors together influence disease risk [8]. • Methods for gene-environment interactions depend on exposure type and genetic architecture [9].
Data Acquisition	• What is the required sample size to have sufficient statistical power for detecting the effects of genetic variation and gene-environment interactions? • Is the dataset diverse in terms of both genetic and social variation? • How is diversity measured in the dataset?	• Sample size may limit detection of gene-environment interactions [10]. • Datasets that encompass diverse demographic histories, social conditions, and genetic diversity enhance the accuracy of risk models [11]. • Measures of structural discrimination and racism, race/ethnicity, and genetic ancestry are not interchangeable [12].
Analysis	• What statistical genetic methods and epidemiological models can account for genetic and social components of population structure and/or clustering? • Are differences in effect sizes across diverse populations attributable to genetic or environmental variation alone or gene-environment interactions? • Can the effects of genomics and the exposome/environment on intermediate -omic variation be disentangled?	• PC-AiR is a method for inferring ancestry that accounts for related samples [13]. • Ecological and multi-level studies have specific fallacies and sources of error [14] • Big data approaches for complex environmental exposures altering biological pathways throughout the life course [15].
Interpretation	• Are the findings generalizable to the entire population or do they vary depending on genetically and/or socially stratified groups? • Can results be misinterpreted as a genetic explanation for racial disparities?	• There are four typologies for categorizing social context and genetic effects [16]. • Human genetic diversity is continuous and cannot be mapped onto socially constructed groupings [17].
Dissemination and Application	• Are the research findings clinically actionable or do they indicate a need for broader population level changes (e.g. policy)? • Does language used in dissemination of research avoid reinforcing or generating social gradients or stereotypes?	• Improve precision for the delivery of population-based strategies with genomics and bioinformatics [18]. • Precise language and population descriptors for reporting on genetic variation [19].

shaping how we view distal causes and the pathways through which they influence biological differences. As such, while we agree with Diez Roux that extensive, well-measured data across domains is needed to carry out the vision of complex systems research (and as we noted in the introduction, much progress has been made on this front), integrating genetics and gene-environment interaction with social explanations of disease may not require entirely new ways of thinking. Principles from genetics and genomics already play a critical role in these frameworks. Leveraging these well-established models and anchoring discussions of genetic variation in epidemiology to these models reduces barriers to interdisciplinary discussion and training. Rather than introducing a separate model that includes genetics in contrast to these models, we provide this resource as a pathway to discuss genetic variation alongside traditional conceptual approaches used to explain social determinants of disease.

Genomics and other -omics as tools for testing Ecosocial theory

Nancy Krieger's Ecosocial theory describes how social and contextual exposures are biologically embodied to result in health disparities [7,24]. Ecosocial theory posits that there are socially patterned pathogenic pathways that are impacted by contextual exposures, rather than being consequences of "innate biology," such as genomics. Krieger and others have used Ecosocial theory as a framework for understanding how social and contextual factors – including racism and discrimination – manifest biologically, beyond disparities caused by proximal factors like access to health care or individual health behaviors. While Ecosocial theory is rooted in social epidemiology principles, genetic epidemiologists have a complementary toolkit that can be used to optimize studies of embodiment. We illustrate two ways the inclusion of genetic and -omic information can improve our understanding of social causes of disease: first by accounting for genetic confounders of the relationship between contextual determinants and health disparities, and second by exploring biological causal mechanisms through which social determinants manifest as health disparities.

Considering genetic influence in Ecosocial theory

Consider a modification to an example from Krieger's seminal *Theories for Social Epidemiology in the 21st century: an Ecosocial Perspective* [7]. Krieger uses Ecosocial theory to highlight that *racism* – not *race*—contributes to disparities in traits like kidney function and blood pressure, specifically stating that Ecosocial theory "recasts alleged 'racial' differences in biology (e.g. kidney function, blood pressure) as mutable and embodied biological expressions of racism [7]." This is a critical distinction, as race itself does not have a biological basis, while structural and interpersonal *racism* can shape access to resources and exposure to stress. However, to identify specific biological pathways of embodiment influenced by racism, it is crucial to consider the role of genetic variation and patterns of genetic ancestry on outcomes of interest. Black Americans are two to four times more likely to develop end-stage renal disease (ESRD) compared to White Americans [25–27]. Much of this disparity is likely driven by the structural and interpersonal racism captured in the Ecosocial model. At the same time, two variants in the *APOL1* gene confer 7 to 10-fold risk for kidney disease and ESRD [28]. These variants are almost exclusively found in those with recent African ancestry, with a recent study reporting that approximately 35% of African Americans carry at least one risk allele [29]. Disparities in kidney disease cannot be entirely explained by genetic variation, but considering these genetic risk factors can help untangle a complicated overlapping relationship between racism, genetic risk, and ancestry. For example, adjusting for *APOL1* genotype when comparing differences in social and contextual determinants would help isolate or refine the direct impacts of racism on health disparities.

Social and genetic risk factors may also act in opposite directions, potentially masking the effects of each when not jointly considered. For example, Hispanic and Black Americans are 1.5–2 times more likely to develop Alzheimer's disease (AD) compared to white Americans [30]. This disparity is primarily due to structural inequalities affecting modifiable risk factors including education, blood pressure, and exposure to toxic pollutants. Concurrently, the *APOEe4* allele, the strongest genetic risk factor for AD, has a lower effect in those with African ancestry compared to those primarily of European descent, likely due to genetic variation nearby the *APOEe4* allele that modifies its effect [31]. One variant found to be protective, rs10423769, is far more common in African populations (allele frequency 0.11) than in European populations (0.001). Ignoring this type of genetic information risks underestimating the impact of social risk factors in AD risk. This example focuses on a single variant, and individuals will have a collection of protective and risk variants, but the point remains that genetic nuances could lead to more precise estimates of the effects of non-genetic risks.

Integrating -omics data to uncover pathways of embodiment

Genetic epidemiologists are increasingly leveraging other -omic data to interpret how genetic variants affect biological pathways underlying disease. This multi-level, high-dimensional approach of integrating genomic, epigenomic, transcriptomic, proteomic, metabolomic and other biological -omic layers is often referred to as a "systems biology" approach [32,33]. Briefly, genome-wide association studies are only able to identify the broad genetic regions within which variation is associated with a trait. With the rise of high-dimensional -omic data, genetic epidemiologists can better pinpoint the specific variants within a region of interest that drive differences in DNA methylation, gene expression, protein and metabolite levels, etc. By synthesizing information from across these biological layers, researchers can better identify underlying biological pathways that mediate genetic associations with a trait. In the past decade, there has also been increasing focus in systems biology to go beyond genetic effects and understand the biological impacts of the human exposome – the accumulation of all environmental exposures one experiences throughout a lifetime, from prenatal development to adulthood [34].

Recall that the Ecosocial theory is focused on "how we literally biologically embody exposures arising from our societal and ecological context [7]," and it becomes apparent that social epidemiology and systems biology approaches have experienced a convergent evolution of sorts, with a shared goal but different language. Genetic epidemiologists have the tools to operationalize questions of how experiences and environmental differences are embodied or "get under the skin." Indeed, recent publications have used the analysis of differentially expressed genes to uncover cellular mechanisms associated with smoking and alcohol behavior [35] and contextual determinants like air pollution levels [36]. Returning to the kidney disease example, using a systems biology approach of investigating differences in expression quantitative trait loci after stratifying by genotype could identify specific patterns of biological change due to differences in the exposome, which could include effects of racism.

Recognizing that Ecosocial theory and systems biology are complementary frameworks can improve collaboration between social and biological scientists, strengthening the public health impact of both fields. Understanding biological mechanisms that perpetuate health disparities can play an important role in achieving health equity. Greater collaboration could uncover how exposures like access to fresh produce or experiences of discrimination contribute to distinct molecular changes at various biological layers. Highlighting significant differences in methylation, protein modifications, metabolomic profiles, etc. among those with the same genotype or genetic risk profiles can have rhetorical power for policy change. Biological differences that exist after controlling for genetic variation underscore the need for upstream structural interventions like policy change by elucidating causal mechanisms that link social context and discrimination to health. Efforts in precision medicine or precision public health can also be strengthened by

including social determinants alongside genetic profiles in order to improve risk stratification strategies used to deliver tailored treatments.

Incorporating genetic variation into tests of Fundamental Cause theory

In addition to pathways of embodiment, genetic variation belongs in discussions of fundamental drivers of health. Link and Phelan's Fundamental Cause Theory (FCT) is a theoretical framework that describes why health inequalities persist despite innovations in disease prevention and treatment [6]. According to FCT, there are certain factors – fundamental causes – that influence health outcomes due to their overarching structural effects, regardless of changes in exposures or systems over time.

There are four features of a fundamental cause: (1) they influence multiple disease outcomes; (2) they affect diseases through multiple risk factors; (3) the association between the fundamental cause and health is reproduced over time; and (4) they influence access to flexible resources like knowledge, connections, and power that can be used to avoid risks or minimize consequences. Proposed fundamental causes of health include socioeconomic status (SES) [6,37], stigma [38], and racism [39].

Since FCT was first introduced in 1995, thousands of studies have been conducted linking socioeconomic status to health and extended the theory to other fundamental causes. In 2021, Clouston and Link conducted a quarter-century retrospective on FCT, synthesizing literature that aims to extend the original theory [40]. They emphasize that beyond meeting the initial criteria posed in "Social conditions as fundamental causes of disease", the strength of FCT relies on persistent evidence of associations between fundamental causes and health outcomes. Based on the synthesis of the literature thus far, Clouston and Link provide a set of approaches related to the influence of flexible resources that can be used to test new associations and predictions made based on FCT. Here, we show how genomic considerations strengthen these approaches and could be considered in new tests of the theory using Clouston and Link's proposed approaches: (1) Disease preventability, (2) Preventability shifts, and (3) Manipulated preventability [40].

Disease preventability and heritability

The *disease preventability* and *preventability shifts* approaches posit that (1) the strength of association between a fundamental cause and disease outcomes should vary depending on how preventable a disease is, with stronger associations between fundamental causes and highly preventable diseases, and (2) shifts in social gradients of disease occur due to the discovery of new prevention strategies; social gradients are stronger after new knowledge is discovered and/or prevention strategies arise). Genetic variation is not mentioned in Clouston and Link's descriptions of these approaches, but it is an underlying element of disease preventability and an important modifier of the relationship between prevention and social gradients. The strength of association between fundamental causes and an outcome depends directly on the trait's genetic architecture and heritability.

Heritability can be defined as the proportion of trait variation that is explained by genetic variation [41]. Complex diseases vary greatly in heritability. For example, schizophrenia has an estimated heritability of up to 80%, while estimates for atherosclerosis and depression range from approximately 50% and 30%, respectively [42,43]. Understanding the heritability of a trait can help identify situations where tests of fundamental cause theory are most appropriate. Diseases that are highly heritable are less likely to be preventable, as their variation is more likely to be influenced by non-modifiable, genetic risk factors rather than flexible resources. These traits will have more subtle associations with fundamental causes, although there are exceptions. Notable exceptions include diseases for which prevention is designed around genetic risk screening, such as heritable breast and ovarian cancer. In these cases, the *preventability shifts* approach becomes more relevant. Indeed, social gradients in cancer outcomes were induced or shifted by unequal access to care following the development of genetic screening [44].

Traits with low heritability, where variation in the outcome is primarily due to non-genetic factors lead to more predictable social gradients, as flexible resources have more influence over these traits. Clouston & Link point to lung cancer outcomes as an example where new public health knowledge (e.g. smoking causes lung cancer) is distributed unevenly, and those with advantaged sets of flexible resources are more likely to benefit from the knowledge [40,45]. Genetic variation plays a small role for this outcome, so inequalities or gradients across socioeconomic status, racism, and other fundamental causes are likely to be more apparent.

However, for more heritable complex diseases caused by a mix of polygenic and environmental factors, the strength of fundamental cause associations will be less consistent, and the effectiveness of prevention strategies will vary depending on one's overall genetic risk. For these types of diseases, many variants throughout the genome each confer a small risk or protective effect that, when added together, equals the overall genetic risk [46]. Under the assumption that genetic risk is normally distributed across the population and genetic and environmental effects are additive, individuals on each extreme of the distribution will have an accumulation of either protective or risk variants such that any changes in environmental (non-genetic) factors is unlikely to impact disease outcome [8]. Diseases that have been assumed to operate under this type of liability threshold model include Type II diabetes, schizophrenia, sporadic cancers, and Alzheimer's disease [47,48].

Alternatively, genetic and environmental effects can be multiplicative rather than additive for certain traits, and changes in modifiable risk factors may have substantially different preventative effects for those who have specific genetic variants. Well-replicated gene-environment interactions include the multiplicative effects of alcohol consumption and *ALDH2* variation on esophageal cancer risk and *NAT2* and smoking on bladder cancer [49–52]. For both additive and multiplicative gene-environment models, obtaining genetic risk information can help identify individuals for whom disease is more highly preventable – allowing for the fine-tuning of tests for fundamental causes. Genomic risk aggregation and prediction based on genome-wide association study results is an area of active research in genetic epidemiology. Considering genetic determinants does not weaken the argument for studying fundamental causes, instead, it can fine-tune our understanding of which diseases are most preventable (and for whom) or explain unexpected and/or attenuated preventability shifts.

Manipulated preventability and systemic inequality in genomic data

The *manipulated preventability approach* for testing fundamental causes can also be enhanced by considering the role of genetics. The manipulated preventability approach is applied in cases where an intervention is formally tested, such as a clinical trial. In these cases, Clouston and Link predict that there will be inequalities in the derived benefit from intervention because those with more advantage sets of flexible resources e.g. individuals of higher SES or those who have experienced less racism will be able to adhere to a protocol more consistently [40]. Manipulated preventability differs from preventability shifts because it posits that beyond social gradients arising due to delayed or restricted access or knowledge, inequalities occur because flexible resources also influence the capability of deriving maximum benefit from an intervention. We propose that by incorporating genetics, it may also be possible to test the prediction that fundamental causes lead to inequalities in the maximum benefit of intervention itself. Genetic variation may be important to consider as a cause of differences in adherence to protocol. Pharmacogenomic considerations present a

prime example.

Genetic variation directly influences the effectiveness of treatments and adverse events. Even if an effective treatment exists, the accessibility and efficacy of the treatment will vary depending on genetic profile. Pharmacogenetic response to the anticoagulant Warfarin is one example: among an otherwise similar group of patients, the genotype in two genes (*CYP2C9* and *VKORC1*) results in the same dosage of Warfarin being effective in some patients and fatal in others [53,54]. Dosing recommendations were largely based on genotypes common in those with European ancestry derived *CYP2C9* and *VKORC1* genes. Warfarin-induced hospitalizations disproportionately affect non-White populations because of the combination of the differences in allele frequencies across ancestry backgrounds combined with disparities in which populations were considered for dosing recommendations [55].

The overwhelming underrepresentation of participants with non-European ancestry in genetic studies cannot be emphasized enough [56]. The biased genetic datasets have led to wide disparities in the utility of genetic findings, and there is a growing literature on the inaccuracy of genetic prediction models for individuals without European ancestry [57,58]. As drug development, clinical trials, and gene-environment studies are downstream applications of genetic research, there is systematic inequality in the benefit or effectiveness of new knowledge. The adverse events in Warfarin present an extreme example of differences in drug efficacy that could be due to genetic variation. Even though many genetic differences in drug efficacy are more subtle, genetic variation can influence perception of the value of clinical intervention and affect decisions to adhere to medical advice and assigned protocols.

In tests using a manipulated preventability approach, researchers should consider the following: those who have genetic variation that is well-represented in genetic studies are poised to derive more benefit from the downstream interventions; if genetic information is not obtained, differences in manipulated preventability tests could be misattributed to social factors.

Applying FCT to genetic epidemiology

Thus far, we have focused on how genetics can be incorporated into FCT – not to downplay the role of social determinants, but to highlight how genetics can support studies traditionally rooted in social epidemiology. However, the consideration of fundamental causes can also improve genetic epidemiology research and the utility of genetics for population health. FCT can help genetic epidemiologists better understand why social determinants are persistently relevant to health outcomes, and to incorporate this knowledge into their study designs and their interpretation of results.

Many genetic epidemiologists are not trained to think critically about the role of genetics within a larger contextual system. As a result, genetic epidemiology studies may undervalue the effects of socioeconomic status or other social determinants of health and overemphasize the role of ancestry-specific genetic variants to explain health disparities. We have provided numerous examples that show how genetic variation may partially explain systematic differences in health outcomes due to the different frequencies of disease-associated variants across ancestral backgrounds, but it is equally important to acknowledge that studying the effects of genetics in isolation will also lead to incomplete understanding of genetic effects. Differences in socioeconomic status or other fundamental causes can lead to substantial differences in choices and behaviors that modify the association between genetic variants and health outcomes.

Genetic datasets have not always prioritized collecting high quality, multi-level environmental and social factors across diverse populations. As genetic datasets continue to grow and we place greater importance on understanding gene-environment interactions, genetic epidemiology would benefit from looking to FCT to ensure diversity of participants across features like socioeconomic status, race, and geographically linked exposures. Social epidemiologists can provide expertise on which exposures are relevant and how to collect reliable and valid measures of those exposures.

FCT also frames why it is important to control for population structure in genetic association studies. It is not merely because there is population specific genetic variation, but because non-genetic fundamental causes including racism, persistently induce systematic gradients in health outcomes. FCT emphasizes that these gradients remain regardless of innovations in public health interventions or scientific knowledge [6]. Because genetic variation is not distributed randomly, and there is correlation between genetic ancestry and social stratification, studies that aim to identify genetic markers associated with disease outcome will pick up spurious hits of population-specific variants that have no effect on disease outcome. Many statistical toolkits have been developed to control for population structure in genome-wide association studies [13,59], but results from these studies must still be followed up carefully to understand whether significant hits are implicating true causal variants. FCT is a useful framework for explaining why genetic associations may have social, rather than biological, explanations.

Conclusion

We have explored the intersection of genetic and social epidemiology using Ecosocial and Fundamental cause theories as guiding frameworks. Specifically, we show that understanding pathways of embodiment – key to Ecosocial theory – depends on the examination of multi-omic data in addition to disentangling the effects of genetic variation and the exposome/contextual determinants throughout the life course. We also demonstrate that considerations of genetic variation can refine tests of fundamental causes, noting that the preventability of disease and shifts in this preventability are dependent on genetic architecture in addition to social determinants.

This is far from the first call for greater integration between genetics and public health. Prior to the current -omic era of big data, Galea et al. and Diez Roux each advocated for complex systems as the path forward in epidemiology [5,60]. Since these early calls, health data and bioinformatics approaches have exploded, yet cross-disciplinary studies of etiology have continued to lag. In fact, there has been ongoing debate over the place of genomics and the quest for 'precision' in public health, with concerns that the focus on genomics diverts resources from more impactful targets of population-level intervention, namely social determinants [61,62]. However, we argue that efforts to understand genetic etiology are not limited to those aimed at genetic intervention. Rather than detracting from population-level interventions, the availability of genomic and other-omic data can aid epidemiologists in evaluating health outcomes and in recommending appropriate policy.

Theories provide a way of seeing, and without frameworks that fully consider the role of genetic variation, social epidemiologists are limited in seeing how genetic data could contribute to their efforts and genetic epidemiologists are limited in seeing how their findings fit into broader efforts to improve public health. Integrating genetics into traditional epidemiological frameworks can facilitate more careful and comprehensive treatment of the role of genetics and non-genetic factors on health, and aid in the design of epidemiological studies. We have described extensions to two social epidemiological theories, arguing for the natural integration of genetics into their principles. Increasing the use and visibility of genetics and genomics in epidemiology does not require entirely new ways of thinking; genetic variation naturally fits in and improves upon existing models. Some applications of the interdisciplinary frameworks can be implemented immediately with existing data, but as the field moves forward with larger biobanks and longitudinal studies, it is important to keep these complex systems conceptual frameworks at the forefront. As more experts and trainees alike learn to think about genetic and contextual perspectives of epidemiology simultaneously, we are bound to discover more opportunities and challenges in integrating across epidemiological fields.

Funding

This publication was supported by the National Institutes of Health [F99AG079792 and K01AG071689].

CRediT authorship contribution statement

Diane Xue: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Anjum Hajat:** Writing – review & editing, Conceptualization. **Alison E. Fohner:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham heart study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014;383:999–1008.
[2] Harris KM, Halpern CT, Whitsel EA, Hussey JM, Killeya-Jones LA, Tabor J, et al. Cohort profile: the National Longitudinal Study of adolescent to adult health (add health). Int J Epidemiol 2019;48. https://doi.org/10.1093/ije/dyz115. 1415–1415k.
[3] The "all of us" research program. N Engl J Med 2019;381:668–76. https://doi.org/10.1056/NEJMsr1809937.
[4] Victora CG, Huttly SR, Fuchs SC, Olinto MT. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. Int J Epidemiol 1997;26:224–7. https://doi.org/10.1093/ije/26.1.224.
[5] Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health 2011;101:1627–34. https://doi.org/10.2105/AJPH.2011.300149.
[6] Link B, Phelan J. Social conditions as fundamental causes of health inequalities. J Health Soc Behav 1995:80–94. https://doi.org/10.2307/j.ctv16h2n9s.4.
[7] Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol 2001;30:668–77. https://doi.org/10.1093/ije/30.4.668.
[8] Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 1965;29:51–76.
[9] Ritz BR, Chatterjee N, Garcia-Closas M, Gauderman WJ, Pierce BL, Kraft P, et al. Lessons learned from past gene-environment interaction successes. Am J Epidemiol 2017;186:778–86.
[10] Murcray CE, Lewinger JP, Conti DV, Thomas DC, Gauderman WJ. Sample size requirements to detect gene-environment interactions in genome-wide association studies. Genet Epidemiol 2011;35:201–10.
[11] Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet 2017;100:635–49.
[12] Adkins-Jackson PB, Chantarat T, Bailey ZD, Ponce NA. Measuring structural racism: a guide for epidemiologists and other health researchers. Am J Epidemiol 2022;191:539–47.
[13] Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol 2015. https://doi.org/10.1002/gepi.21896.
[14] Blakely TA, Woodward AJ. Ecological effects in multi-level studies. J Epidemiol Community Health 1978;2000(54):367–74.
[15] Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 2005;14:1847–50.
[16] Shanahan MJ, Hofer SM. Social context in gene–environment interactions: retrospect and prospect. J Gerontol B Psychol Sci Soc Sci 2005;60:65–76.
[17] Lewontin RC. The apportionment of human diversity. In: Dobzhansky T, Hecht MK, Steere WC, editors. Evolutionary biology. Vol. 6. New York, NY: Springer US; 1972. p. 381–98. https://doi.org/10.1007/978-1-4684-9063-3_14.
[18] Roberts MC, Fohner AE, Landry L, Olstad DL, Smit AK, Turbitt E, et al. Advancing precision public health using human genomics: examples from the field and future research opportunities. Genome Med 2021;13:97.
[19] National Academies of Sciences and Medicine E. Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field. 2023.
[20] Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc Sci Med 2006;62:1650–71.
[21] Boardman JD, Daw J, Freese J. Defining the environment in Gene–Environment research: Lessons from social epidemiology. Am J Public Health 2013;103. https://doi.org/10.2105/AJPH.2013.301355. S64–72.
[22] Shantz E, Elliott SJ. From social determinants to social epigenetics: health geographies of chronic disease. Health Place 2021;69:102561. https://doi.org/10.1016/j.healthplace.2021.102561.
[23] Diez Roux AV. Conceptual approaches to the study of health disparities. Annu Rev Public Health 2012;33:41–58.
[24] Krieger N. Embodiment: a conceptual glossary for epidemiology. J Epidemiol Community Health 1978;2005(59):350. https://doi.org/10.1136/jech.2004.024562.
[25] Norton JM, Moxey-Mims MM, Eggers PW, Narva AS, Star RA, Kimmel PL, et al. Social determinants of racial disparities in CKD. J Am Soc Nephrol 2016;27. https://doi.org/10.1681/ASN.2016010027. 2576 LP – 2595.
[26] Derose SF, Rutkowski MP, Crooks PW, Shi JM, Wang JQ, Kalantar-Zadeh K, et al. Racial differences in estimated GFR decline, ESRD, and mortality in an integrated health system. Am J Kidney Dis 2013;62:236–44.
[27] Bick AG, Akwo E, Robinson-Cohen C, Lee K, Lynch J, Assimes TL, et al. Association of APOL1 risk alleles with cardiovascular disease in blacks in the million veteran program. Circulation 2019;140:1031–40. https://doi.org/10.1161/CIRCULATIONAHA.118.036589.
[28] Foster MC, Coresh J, Fornage M, Astor BC, Grams M, Franceschini N, et al. APOL1 variants associate with increased risk of CKD among African Americans. J Am Soc Nephrol 2013;24. https://doi.org/10.1681/ASN.2013010113. 1484 LP – 1491.
[29] Daneshpajouhnejad P, Kopp JB, Winkler CA, Rosenberg AZ. The evolving story of apolipoprotein L1 nephropathy: the end of the beginning. Nat Rev Nephrol 2022;18:307–20. https://doi.org/10.1038/s41581-022-00538-3.
[30] Alzheimer''s Association. Alzheimer's disease facts and figures. Alzheimers Dement 2023;2023(19):1598–695. https://doi.org/10.1002/alz.13016.
[31] Rajabli F, Beecham GW, Hendrie HC, Baiyewu O, Ogunniyi A, Gao S, et al. A locus at 19q13.31 significantly reduces the ApoE ε4 risk for Alzheimer's disease in African Ancestry. PLoS Genet 2022;18:e1009977.
[32] Haas R, Zelezniak A, Iacovacci J, Kamrad S, Townsend S, Ralser M. Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. Curr Opin Syst Biol 2017;6:37–45. https://doi.org/10.1016/j.coisb.2017.08.009.
[33] Sun YV, Hu Y-J. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. Adv Genet 2016;93:147–90.
[34] Vermeulen R, Schymanski EL, Barabási A-L, Miller GW. The exposome and health: where chemistry meets biology. Science 1979;2020(367):392–6.
[35] Slattery ML, Pellatt DF, Mullany LE, Wolff RK. Differential gene expression in Colon tissue associated with diet, lifestyle, and related oxidative stress. PLoS One 2015;10:e0134406.
[36] Chu J, Hart JE, Chhabra D, Garshick E, Raby BA, Laden F. Gene expression network analyses in response to air pollution exposures in the trucking industry. Environ Health 2016;15:101. https://doi.org/10.1186/s12940-016-0187-z.
[37] Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav 2010;51:S28–40.
[38] Hatzenbuehler ML, Phelan JC, Link BG. Stigma as a fundamental cause of population health inequalities. Am J Public Health 2013;103:813–21.
[39] Phelan JC, Link BG. Is racism a fundamental cause of inequalities in health? Annu Rev Sociol 2015;41:311–30.
[40] Clouston SAP, Link BG. A retrospective on fundamental cause theory: state of the literature and goals for the future. Annu Rev Sociol 2021;47:131–56.
[41] Mayhew J, A, Meyre D.. Assessing the heritability of complex traits in humans: methodological challenges and opportunities. Curr Genomics 2017;18:332–40.
[42] Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 2000;157:1552–62.
[43] Dai X, Wiernek S, Evans JP, Runge MS. Genetics of coronary artery disease and myocardial infarction. World J Cardiol 2016;8:1.
[44] Zapata-Moya ÁR, Willems B, Bracke P. The (re) production of health inequalities through the process of disseminating preventive innovations: the dynamic influence of socioeconomic status. Health Sociol Rev 2019;28:177–93.
[45] Rubin MS, Clouston S, Link BG. A fundamental cause approach to the study of disparities in lung cancer and pancreatic cancer mortality in the United States. Soc Sci Med 2014;100:54–61.
[46] Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. Cell 2017;169:1177–86.
[47] Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. Eur J Hum Genet 2010;18:1039–43. https://doi.org/10.1038/ejhg.2009.177.
[48] Dahlqwist E, Magnusson PKE, Pawitan Y, Sjölander A. On the relationship between the heritability and the attributable fraction. Hum Genet 2019;138:425–35.
[49] Vineis P, Marinelli D, Autrup H, Brockmoller J, Cascorbi I, Daly AK, et al. Current smoking, occupation, N-Acetyltransferase-2 and bladder cancer: A pooled analysis of genotype-based studies. Cancer Epidemiol Biomarkers Prev 2001;10:1249–52.
[50] Song Y, Qi X, Liu X. N-acetyltransferase 2 polymorphism is associated with bladder cancer risk: an updated meta-analysis based on 54 case-control studies. Gene 2020;757:144924. https://doi.org/10.1016/j.gene.2020.144924.
[51] Teleka S, Jochems SHJ, Jirström K, Stocks T. The interaction between smoking and bladder cancer genetic variants on urothelial cancer risk by disease aggressiveness. Cancer Med 2022;11:2896–905. https://doi.org/10.1002/cam4.4654.
[52] Matsuo K, Hamajima N, Shinoda M, Hatooka S, Inoue M, Takezaki T, et al. Gene–environment interaction between an aldehyde dehydrogenase-2 (ALDH2) polymorphism and alcohol consumption for the risk of esophageal cancer. Carcinogenesis 2001;22:913–6. https://doi.org/10.1093/carcin/22.6.913.
[53] Lee MTM, Klein TE. Pharmacogenetics of warfarin: challenges and opportunities. J Hum Genet 2013;58:334–8. https://doi.org/10.1038/jhg.2013.40.
[54] Shendre A, Dillon C, Limdi NA. Pharmacogenetics of warfarin dosing in patients of African and European ancestry. Pharmacogenomics 2018;19:1357–71. https://doi.org/10.2217/pgs-2018-0146.

[55] Yuh-Jer SA, YJ F, BS S, JM B, Wansu C. Racial/ethnic differences in the risk of intracranial hemorrhage among patients with atrial fibrillation. J Am Coll Cardiol 2007;50:309–15. https://doi.org/10.1016/j.jacc.2007.01.098.
[56] Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature 2016;538: 161–4. https://doi.org/10.1038/538161a.
[57] Privé F, Aschard H, Carmi S, Folkersen L, Hoggart C, O'Reilly PF, et al. Portability of 245 polygenic scores when derived from the UK biobank and applied to 9 ancestry groups from the same cohort. Am J Hum Genet 2022;109:12–23. https://doi.org/10.1016/j.ajhg.2021.11.008.
[58] Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019; 51:584–91. https://doi.org/10.1038/s41588-019-0379-x.
[59] Sul JH, Martin LS, Eskin E. Population structure in genetic studies: confounding factors and mixed models. PLoS Genet 2018;14:e1007309.
[60] Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol 2010;39:97–106.
[61] Taylor-Robinson D, Kee F. Precision public health—the Emperor's new clothes. Int J Epidemiol 2019;48:1–6. https://doi.org/10.1093/ije/dyy184.
[62] Merlin C, Ronald B, Sandro G. "Precision" public health — between novelty and hype. N Engl J Med 2018;379:1398–400. https://doi.org/10.1056/NEJMp1806634.