CHAPTER

4

Genomic data measures and methods: a primer for social scientists

Erin B. Ware[1] *and Jessica D. Faul*[2]

[1]Population, Neurodevelopment, and Genetics, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, United States [2]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, United States

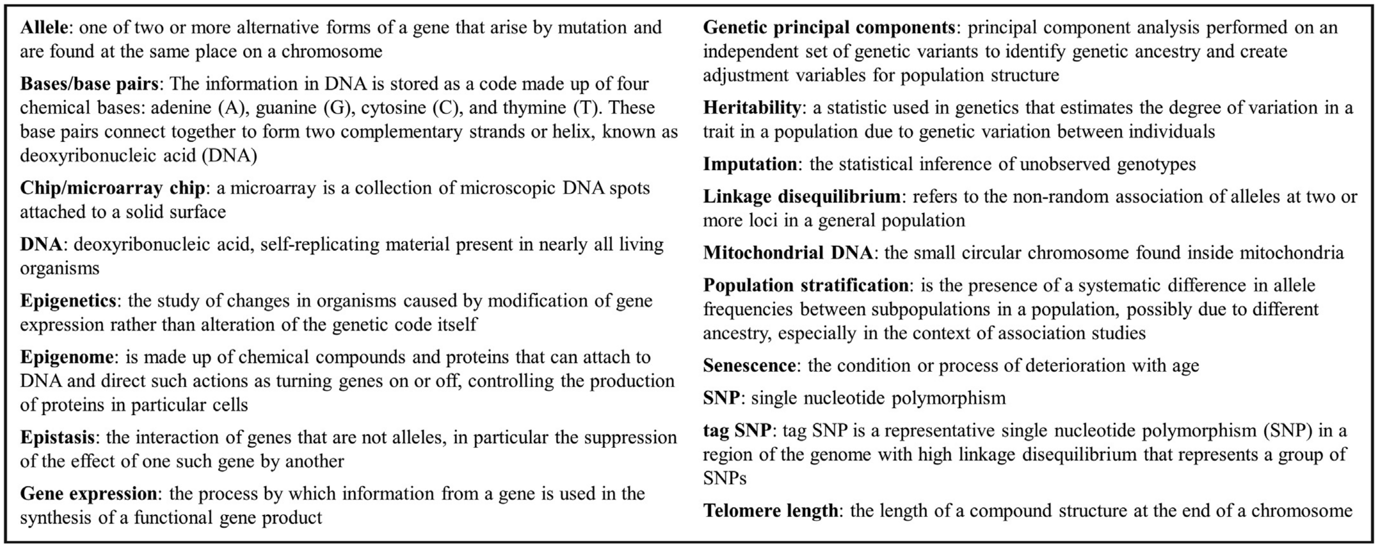
OUTLINE

Introduction

The process of biological aging refers to the decline in the function and structures of an organism. Molecular and cellular modifications control some amount of this decline, which may have various effects at the individual-level across the life course. Physiological changes that occur over the life course due to molecular, cellular, and tissue damage can decrease the capacity to maintain homeostasis—or a state of equilibrium within the body—in certain conditions, and may increase risk for many diseases, such as cardiovascular, cancer, and neurodegenerative disorders, and even mortality (Fraga, Agrelo, & Esteller, 2007; Fraga & Esteller, 2007). These modifications or physiological changes arise from genetic and epigenetic interactions that may depend on hereditary, environmental, and/or stochastic factors. Identifying these factors is complicated by the complexity of the aging process and by heterogeneity across individuals and even within tissue types within an individual. Cellular aging—called senescence—is a result of both internal and external aging factors, due to both the gradual accumulation of DNA damage and epigenetic changes in DNA structure that affect gene expression and may lead to altered cell function (Rodríguez-Rodero et al., 2011) (see Fig. 4.1 for genetic terms). The multifactorial nature of aging makes disentangling purely genetic factors from environmental factors difficult. While the genotype determines the variation in life span among species or individuals, this variation is in part a result of the differential accumulation of molecular errors like DNA damage and epigenetic changes across different socially defined groups.

Over the past two decades, the field of genetics has undergone many technological changes. As a direct result, the type, quality, and time taken to generate and analyze these data has changed drastically. We will briefly outline some of the established statistical genetics

Handbook of Aging and the Social Sciences.
DOI: https://doi.org/10.1016/B978-0-12-815970-5.00004-8



Allele: one of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome

Bases/base pairs: The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). These base pairs connect together to form two complementary strands or helix, known as deoxyribonucleic acid (DNA)

Chip/microarray chip: a microarray is a collection of microscopic DNA spots attached to a solid surface

DNA: deoxyribonucleic acid, self-replicating material present in nearly all living organisms

Epigenetics: the study of changes in organisms caused by modification of gene expression rather than alteration of the genetic code itself

Epigenome: is made up of chemical compounds and proteins that can attach to DNA and direct such actions as turning genes on or off, controlling the production of proteins in particular cells

Epistasis: the interaction of genes that are not alleles, in particular the suppression of the effect of one such gene by another

Gene expression: the process by which information from a gene is used in the synthesis of a functional gene product

Genetic principal components: principal component analysis performed on an independent set of genetic variants to identify genetic ancestry and create adjustment variables for population structure

Heritability: a statistic used in genetics that estimates the degree of variation in a trait in a population due to genetic variation between individuals

Imputation: the statistical inference of unobserved genotypes

Linkage disequilibrium: refers to the non-random association of alleles at two or more loci in a general population

Mitochondrial DNA: the small circular chromosome found inside mitochondria

Population stratification: is the presence of a systematic difference in allele frequencies between subpopulations in a population, possibly due to different ancestry, especially in the context of association studies

Senescence: the condition or process of deterioration with age

SNP: single nucleotide polymorphism

tag SNP: tag SNP is a representative single nucleotide polymorphism (SNP) in a region of the genome with high linkage disequilibrium that represents a group of SNPs

Telomere length: the length of a compound structure at the end of a chromosome

FIGURE 4.1 Genetic terms defined.

methodologies that have helped the field become what it is today. We will also highlight data generation, methods by which these data are being analyzed, as well as important social, epidemiological, and statistical considerations necessary for proper data analysis in aging populations. Healthy aging and longevity in humans is modulated by a combination of genomic and nongenomic factors. Why are some individuals at increased risk for age-related cognitive decline and eventual dementia? Can we increase the human life span through interventions like calorie restriction or treatment with metformin? A portion of the answers to these questions resides with understanding the genetic and molecular basis of aging and to what extent traits and outcomes are modulated by genetic background, the environment, and lifestyle factors.

Heritability

Determining whether a trait (be it disease, personality, behavioral) is at its root genetic starts with determining its heritability, that is the overall proportion of the variability of a trait that can be attributed to genetic factors. Though some disease traits are monogenic—caused by a single genetic variant—most traits are complex and stem from interactions between the environment and several, if not many, genetic factors. Different types of studies are used to estimate heritability including twin studies, family studies, and adoption studies (Clinical and Translational Science—2nd ed., 2016). For example, a twin study found estimates for heritability of longevity at 0.26 for males and 0.23 for females (Herskind et al., 1996; Willcox, Willcox, He, Curb, & Suzuki, 2006). Heritability is not constant and can change over time because the variation due to environmental factors may change, correlations between genes and the environment may change, and the variance in genetic values may change. Often, traits become less heritable in later life due to the accumulation of environmental exposures over the life course resulting in traits that may once have been highly heritable but are later confounded by environment (Visscher, Hill, & Wray, 2008).

Genetics

In this chapter, we will explore several of the many kinds of genomic data available to researchers, including genotyped DNA (generally referred to as "genetic" data), DNA methylation, and telomeres. This section will describe genetic data, how it is analyzed, and social, statistical, and epidemiological considerations when using these data in social science research on aging.

Genome-wide data

Genome-wide data collected from tissue samples—usually saliva or blood—reveal the chemical building blocks, or bases, that make up DNA molecules on each chromosome. Table 4.1 shows a list of genotype data available from selected life course and aging studies. A read of a DNA strand will tell you which of four chemical bases is at each position and will also tell you which bases are on the companion strand, since adenine (A) always pairs with thymine (T); and cytosine (C) always pairs with guanine (G). The human genome contains approximately 3 billion of these base pairs, which contain the instructions for making and maintaining life. Humans are 99.9% identical across these 3 billion base pairs (National Human Genome Research Institute, 2015, 2018), but the remaining 0.1% can hold

TABLE 4.1 Genetic data available from life course and aging studies.

Dataset	Study description	Genotype	Epigenetic	Telomere	Weblink
HRS	The Health and Retirement Study (HRS) is a longitudinal project sponsored by the National Institute on Aging (NIA U01AG009740) and the Social Security Administration	X	X	X	https://hrs.isr.umich.edu/data-products/genetic-data
MIDUS	The first national survey of Midlife Development in the US (MIDUS) was conducted in 1995/96 by the MacArthur Foundation Research Network on Successful Midlife Development	X			http://midus.wisc.edu/midus_restricted_data.php
WLS	The Wisconsin Longitudinal Study (WLS) is a long-term study of a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957	X			https://www.ssc.wisc.edu/wlsresearch/data/Request_Genetic_Data_28_June_2017.pdf
AddHealth	The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7–12 in the United States during the 1994–95 school year	X			https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id = phs001367.v1.p1
NICOLA	In 2012, NICOLA was set up to explore why and how certain social, economic and biological factors are changing the lives of older people; to understand how health, lifestyle, financial circumstances and well-being change with age; to understand what it is like to grow older in Ireland	X	X		https://www.qub.ac.uk/sites/NICOLA/FileStore/Filetoupload,844535,en.pdf
TILDA	The Irish LongituDinal Study on Ageing (TILDA) is a large-scale, nationally representative, longitudinal study on aging in Ireland, the overarching aim of which is to make Ireland the best place in the world to grow old	X	X		https://tilda.tcd.ie/data/accessing-data/
ELSA	The English Longitudinal Study of Ageing (ELSA) collects data from people aged over 50 to understand all aspects of aging in England	X			https://www.ebi.ac.uk/ega/studies/EGAS00001001036
MHAS	The Mexican Health and Aging Study (MHAS) is a national longitudinal study of adults 50 years and older in Mexico	In progress			http://www.mhasweb.org/index.aspx
MESA	The Multi-Ethnic Study of Atherosclerosis (MESA) is a medical research study involving more than 6000 men and women from six communities in the United States	X	X	X	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id = phs000420.v6.p3
Lothian	The Lothian Birth Cohorts of 1921 and 1936 are follow-up studies of the Scottish Mental Surveys of 1932 and 1947	X	X	X	https://www.lothianbirthcohort.ed.ac.uk/
ARIC	The Atherosclerosis Risk in Communities Study (ARIC), sponsored by the National Heart, Lung, and Blood Institute (NHLBI) is a prospective epidemiologic study conducted in four US communities	X	X	X	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id = phs000090.v1.p1
FHS	The Framingham Heart Study has been committed to identifying the common factors or characteristics that contribute to cardiovascular disease (CVD)	X	X	X	https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id = phs000307.v15.p11
NHATS	Begun in 2011, the National Health and Aging Trends Study (NHATS) fosters research to guide efforts to	X			https://www.nhatsdata.org/

(*Continued*)

TABLE 4.1 (Continued)

Dataset	Study description	Genotype	Epigenetic	Telomere	Weblink
	reduce disability, maximize health and independent functioning, and enhance quality of life at older ages				
BLSA	The National Institute on Aging's Baltimore Longitudinal Study of Aging (BLSA) answers critical questions about what happens as people get older	X	X		https://blsa.nia.nih.gov/how-apply
InCHIANTI	The InCHIANTI Study (Invecchiare in Chianti, aging in the Chianti area) is currently supported by a grant from the National Institute on Aging with the goal to translate epidemiological research into geriatric clinical tools that makes possible more precise diagnosis and more effective treatment in older persons with mobility problems	X			http://inchiantistudy.net/wp/inchianti-dataset/

AddHealth, National Longitudinal Study of Adolescent to Adult Health; *ARIC*, Atherosclerosis Risk in Communities Study; *BLSA*, Baltimore Longitudinal Study of Aging; *ELSA*, The English Longitudinal Study of Ageing; *FHS*, Framingham Heart Study; *HRS*, Health and Retirement Study; *InCHIANTI*, Aging in the Chianti Area; *Lothian*, Lothian Birth Cohorts of 1921 and 1936; *MESA*, Multi-Ethnic Study of Atherosclerosis; *MHAS*, Mexican Health & Aging Study; *MIDUS*, Midlife in the United States; *NHATS*, National health and Aging Trends Study; *NICOLA*, Northern Ireland Cohort for the Longitudinal Study of Ageing; *TILDA*, The Irish Longitudinal Study on Ageing; *WLS*, Wisconsin Longitudinal Study.

important keys to determining the origins of traits. Single nucleotide polymorphisms, or SNPs, are locations on the genome and vary across populations.

Reading genome-wide data is done through commercially available chips, which recognize certain positions in the genome. When a base varies at a SNP, the different options for the base are alleles. Remember that the typical individual has *two* copies of each chromosome and therefore at any given "location" (i.e., chromosome 1, position 100000001) there are two possible pairs of bases per individual [see Fig. 4.2, Panel (A)]. It is likely that these alleles do not match in the intuitive AT CG pattern because, though they are at the same position in an individual, they are on different chromosomes and do not need to chemically bond to one another. At a SNP, alleles can be minor or major based on their frequency in a given population. Minor alleles have a prevalence of <50% in a given population. The definition of a minor or major allele must occur at the population level and these frequencies are ancestry specific. When we summarize genome-wide data in a study, we count the number of a certain allele an individual has, creating a 0, 1, or 2 value for each site, for each individual. This allele might be the minor allele, or major allele, or simply the first alphabetical allele depending on which allele is summed. Genotyping chips contain varying numbers of SNPs (typically from 500,000 to several million) where some of these SNPs are tag SNPs meant to represent a block of correlated SNPs called a linkage disequilibrium (LD) block.

Genetic ancestry is a prediction of biogeographical origin based on sections of the genome that are unique to specific groups. Population stratification represents systematic differences in allele frequency between subpopulations in a population—sometimes due to differences in ancestry. Since there are known correlation structures in the frequencies of alleles by ancestry (LD patterns), researchers could take a genotyping chip with a limited number of tag SNPs and amplify it to a much larger set of SNPs by comparing with a reference panel of fully sequenced individuals of the same ancestry. Amplifying a set of genotype data to a larger set of predicted genotypes using LD is referred to as genotype imputation. Many different imputation panels exist based on homogenous groups of individuals with known ancestries, including the International HapMap Project, the 1000 Genomes Project (1000G) (Genomes Project et al., 2010), the UK10K Project, the Haplotype Reference Consortium (HRC) (McCarthy et al., 2016), and the Trans-Omics for Precision Medicine (TOPMed) program (National Heart, Lung and Blood Institute, 2018). Using imputation can increase the number of SNPs from a genotyping chip from hundreds of thousands up to multiple millions. Researchers can impute SNPs with a known amount of statistical certainty based on an individual's ancestral composition. Using this information, studies or consortia can remove SNPs with low imputation quality. To harmonize genotype data across studies, consortia typically require studies to impute to a specific panel.

Dimension reduction techniques, such as principal components and multidimensional scaling, can take the patterns in these ancestry-linked variants and calculate genetic principal components. These components can account for spurious correlations between genotypes and phenotypes directly related to ancestry composition.

Analysis of genome-wide data

Genome-wide association studies

Genome-wide data are commonly analyzed through a genome-wide association study (GWAS). GWAS

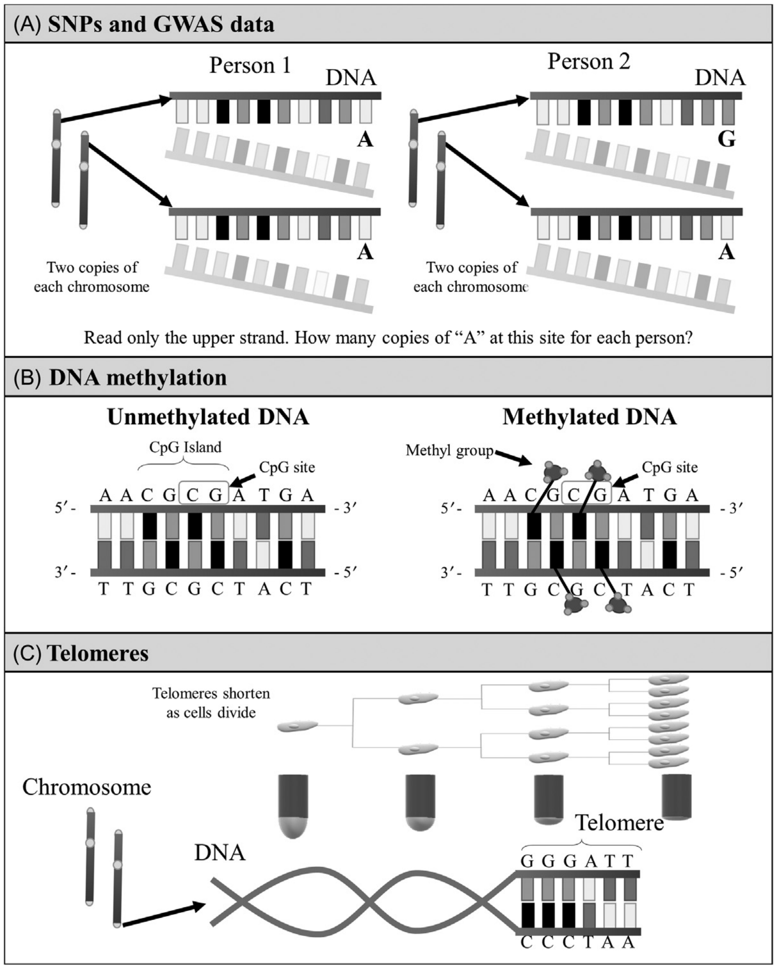


FIGURE 4.2 Heuristic model of three common types of genomic data: (A) genotype data. Panel 1 depicts two individuals, each with two copies of a chromosome, and a short section of DNA. Each colored box represents a nucleotide (adenine, cytosine, thymine, and guanine) with the last nucleotide showing variation across the individuals. Person 1 has an "AA" genotype at this location while Person 2 has a "GA" genotype; (B) DNA methylation data. Panel 2 represents DNA methylation by showing the same segment of DNA from two different cells within the same person; and (C) telomere length data. Panel 3 shows how telomeres—the cap at the end of a chromosome—tend to shorten over the course of cell division. *GWAS*, genome-wide association study; *SNP*, single nucleotide polymorphism.

regress a phenotype on a SNP, for every SNP in the sample. This is equivalent to running hundreds of thousands to millions of linear or logistic regression models. Covariates such as age and sex are often included in these models, as well as genetic principal components representing the broad ancestral composition of the sample. These models are typically run using genetic analysis software (e.g., PLINK, SNPTEST) under the assumption that each SNP has an additive effect on the outcome (Marchini & Band, 2019). It is possible to use more complex modeling strategies (e.g., linear mixed models, generalized estimating equations) and different assumptions of the effect of the SNP on an outcome (e.g., dominant, recessive); however, this sometimes requires writing one's own programs to do so. Due to the large number of tests performed in a GWAS, reliable estimates need large sample sizes.

We consider a variant to be associated with the outcome if certain genetic variants are more frequent in affected individuals than nonaffected. Because of the large number of tests performed, correcting for multiple testing using a more stringent genome-wide significance level of 5×10^{-8} or a suggestive level of 1×10^{-6} is necessary. GWAS are largely hypothesis-generating and point to regions for future study. False discovery rates (FDR) are also a popular method to correct for multiple testing in GWAS (Benjamini & Hochberg, 1995). After a GWAS is performed, results are presented in a Manhattan plot where the x-axis is represented by regions of the genome and the y-axis is the log10(*P*-value) for each SNP. Quantile−quantile plots of the *P*-values plotted against an expected uniform distribution identify large deviations from expectation, indicating a trait is highly polygenic. GWAS are useful in finding genetic variations that contribute to common, complex phenotypes, such as asthma, heart disease, mental illness, cognitive function, Alzheimer's disease, and longevity (National Human Genome Research Institute, 2015, 2018). Researchers can use information from GWAS as pointers to regions of the human genome where phenotype-related variants may reside. For example, in a GWAS of Alzheimer's disease, variants within the APOE region were highly significantly associated with Alzheimer's disease; however, not the variants that create the APOE-ε4/ ε4 genotype (Jansen et al., 2019). Significant GWAS variants may not themselves directly cause the phenotype, and researchers generally need to take further steps, such as sequencing the area identified in the GWAS, to identify the actual causal variants. These steps may help to better identify vulnerable populations and perhaps even target interventions.

Polygenic scores

Polygenic scores (PGS) capitalize on the polygenic nature of most complex phenotypes. PGSs are created by taking population estimates of SNP effects on a phenotype from GWAS and using those values to weight individual genotypes before they are summed (Dudbridge, 2013). PGSs can use all variants that overlap between a GWAS and the set of SNPs in a study (genome-wide scores), a subset of either highly significant variants (sometimes known as "top hits" scores), or a subset of variants selected based on their independent effects after pruning out variants that are in LD (i.e., those that are highly correlated). PGSs do not implicate biological mechanisms, but rather capture a statistical predisposition to a phenotype—and often other, related phenotypes. For example, a PGS for educational attainment has been shown to contain a cognitive and noncognitive component (Lee et al., 2018; Okbay et al., 2016). Information from PGS can also help to quantify genetic risk for diseases like coronary artery disease and to help better estimate the impact of behavioral or environmental modification across risk groups (Khera et al., 2016).

Gene-region analyses

Gene-region analyses are joint tests performed within a specified gene region or SNP sets. The joint effects of genetic variation influence many complex diseases. A large number of group-wise association tests have been developed recently to evaluate SNP sets and their joint association with disease (Han & Pan, 2010; Ionita-Laza, Buxbaum, Laird, & Lange, 2011; Ionita-Laza, Lee, Makarov, Buxbaum, & Lin, 2013; Lee, Emond, et al., 2012; Lee, Wu, & Lin, 2012; Madsen & Browning, 2009; Neale et al., 2011; Price et al., 2010; Tzeng et al., 2011; Wu et al., 2011). Group-wise testing has been shown to alleviate problems with intensive computation and multiple testing as well as lead to more stable results and more biologically relevant interpretations (Beyene, Tritchler, Asimit, & Hamid, 2009; Buil et al., 2009; Qiao et al., 2009).

In particular, principal component-based approaches (PCA), burden score, and variance-component testing have all been proposed as methods to evaluate the joint effect of SNPs on a disease (Chen, Wang, Smith, & Zhang, 2008; Gao et al., 2011; Gauderman, Murcray, Gilliland, & Conti, 2007; Ionita-Laza et al., 2011, 2013; Lee, Wu, et al., 2012). Each of these approaches has its own methodological strengths and limitations but all require careful selection and definition of gene regions/SNP sets, whether gene regions are selected from genome-wide significant GWAS hits, proposed biological pathways, or some other method. Considering the size of LD blocks (regions in the genome with high correlation) in the given population, the inclusion of promoter or enhancer regions around the gene, the number of variants included in the region and at what allele frequency, or additional flanking regions around the start and stop positions of genes can all have implications on the interpretation and inference of the findings.

Gene set enrichment analysis and pathway analysis

Gene set enrichment analysis is a computational method that can determine whether a set of genes (e.g., genes identified from significant variants in a GWAS) disproportionately overlap with other groups of genes whose products are known or predicted to participate in a common biological process (Pers et al., 2015). This method compares the given set of genes to extensive biological annotation databases. Gene set analysis allows us to make statements ascribing an experimental result to changes in underlying gene-based functions.

Pathway analysis, though more common in gene expression analysis, is a promising avenue for SNP data analysis. These types of analyses highlight interactive evaluations of the possible effects of variations on function, regulation, or interaction of gene products (Cirillo, Parnell, & Evelo, 2017). Pathway analysis can help elucidate biological mechanisms by accounting for the polygenic nature of complex diseases while visualizing epistatic (i.e., epistasis, or gene−gene interaction) effects.

Epigenetics

Genes can switch between active phases, producing proteins, and silent, or inactive, phases. These patterns of activation and silencing exist across all genes in a cell, and differentially across cell and tissue types. These patterns of activation are known as the epigenome and the study of these patterns is called epigenetics (Lamb, 2007). Changes to the epigenome do not alter the gene's DNA sequence, but rather its activity.

Epigenetic data

The most commonly studied type of epigenetic change, DNA methylation (DNAm), consists of methyl tags that attach to regions of DNA where a guanine nucleotide follows a cytosine in the linear sequence of bases, or "CpG sites." (Fig. 4.2B). In humans, 70%−80%

of CpG cytosines are methylated (Jabbari & Bernardi, 2004). CpG sites occur with varying frequency across the genome. Areas with a high frequency of CpG sites are called CpG islands (or CG islands). DNA methylation represents just one of the types of epigenetic mechanisms cells use to turn genes "on" and "off." Commercially available microarrays measure DNA methylation. These arrays have probes that cling to and mark the intensity of methylation at a given site in the genome. A computer reads and translates the intensities into methylation beta values representing the ratio of intensities between methylated and unmethylated alleles. These beta values are always between 0 and 1, with 0 representing unmethylated and 1 representing fully methylated sites (Du et al., 2010). While the beta value has a more intuitive biological interpretation, some researchers log transform the beta values producing "M values" to better satisfy statistical modeling assumptions.

Other types of epigenetic changes include histone modifications and posttranslational modifications of amino acids on the amino-terminal tail of histones (Dupont, Armant, & Brenner, 2009). Studying the patterns of gene activation help elucidate how a gene functions under normal conditions, and also how improper activation or inactivation can lead to disorders like obesity and diabetes. Identifying what triggers gene activation also can reveal sensitive periods in development during which cells are susceptible to environmental influences (Naumova et al., 2019).

Analysis of epigenetic data

Epigenome-wide association studies

Epigenetics is of tremendous interest to social scientists because of the interplay between the environment and the epigenome. The environment, exposures like pollution, and diet, can alter the epigenome and subsequently change the activity of genes. Evidence suggests that exposures such as stress and poverty may also be associated with epigenetic alterations (reviewed in Cunliffe, 2016; Notterman & Mitchell, 2015). Epigenetic modification over the life span helps to explain why monozygotic twins, who are genetically identical, deviate phenotypically over time (Bell & Spector, 2011; Webster et al., 2018). Just as GWAS are a common analytic tool in the field of genetic epidemiology, epigenome-wide association studies (EWAS) using DNA methylation data, most commonly from peripheral blood or saliva, are similarly used to identify the common normal variation in the DNA methylome and better understand the molecular basis for disease risk (Flanagan, 2015). While genetic risk of disease is unmodifiable, epigenetic risk may be reversible and/or modifiable. Animal models are providing potential targets for intervention including reducing environmental toxicants and early-life stress, and modifying diet (Phillips & Roth, 2019).

Epigenetic clocks

In general, DNA methylation changes rapidly after birth and during childhood and slows with age, however, in a somewhat predictable way (Vaiserman, 2018). Biomarkers of aging have thus been created from DNA methylation data and have enabled accurate age estimates from DNA from any tissue across the entire life course (Horvath & Raj, 2018). "Epigenetic clocks" connect developmental and cellular maintenance processes to biological age. The development of these scores has allowed researchers to investigate biological age and its relation to chronological age from the methylation profiles at select CpG sites.

A large number of epigenetic clocks exist. Horvath's (2013) original epigenetic clock uses the methylation at 353 CpG sites and is a highly heritable measure for age acceleration across several tissue types (Horvath, 2013). The difference between the predicted age from the DNA methylation profile and chronological age is the average age acceleration. An alternative clock was proposed by Hannum et al (2013) using 71 methylation markers (Hannum et al., 2013). A test of these two clocks conducted with four cohorts found that Horvath's clock predicts values that skew lower than the chronological age in two of the cohorts (Marioni et al., 2015). Hannum's clock, on the other hand, predicts higher chronological age values on all four cohorts. Both had a positive association between age acceleration and mortality. Age acceleration has been found to be associated with diseases in older age including higher cancer and cardiovascular disease-related mortality (Perna et al., 2016; Teschendorff et al., 2010; Zheng et al., 2016) as well as other phenotypes such as frailty and Parkinson's disease (Breitling et al., 2016; Gale, Marioni, Harris, Starr, & Deary, 2018; Horvath & Ritz, 2015). Another clock, termed the skin and blood clock, includes 391 CpG sites that were selected from the regression of chronological age on methylation states of CpGs from a variety of cell types (Horvath et al., 2018). It outperforms both the original Horvath and Hannum methods on accurate age estimates for blood methylation data. An additional clock, DNAm PhenoAge, is an epigenetic biomarker of aging that was constructed from clinical measures of phenotypic age and their relation to novel CpG sites (Levine et al., 2018). This model strongly outperforms other models in relation to predictions for various aging outcomes and may more accurately predict "biological" aging. DNAm PhenoAge is a better predictor of 10- and 20-year survival than the clocks trained on chronological age. Age-adjusted DNAm PhenoAge is also associated with neuropathological hallmarks of Alzheimer's disease. These types of DNAm summary measures may help us better understand the characteristic interindividual variability in age-associated functional decline and disease as well as health disparities (Field et al., 2018).

As methods for collecting and analyzing DNA methylation become more widespread, it is increasingly important to pay attention to differences by tissue and cell type and whether those confound the relationship between DNA methylation, environment, and health. DNA methylation varies by cell type and thus the cell composition of any sample. Most large-scale epidemiologic or cohort studies that measure DNA methylation do so on complex tissue samples (blood, saliva, brain) consisting of many different cell types that have unique epigenetic profiles. As a result, associations between epigenetic differences or change over time and exposures/outcomes are difficult to disentangle from differences or changes in cell type distributions. Some of the early associations between chronological age and DNA methylation in blood were subsequently correctly attributed to age-related changes in cell composition (Jaffe and Irizarry, 2014). Measuring cell composition heterogeneity is critical in understanding epigenetic change and EWAS.

Telomeres

Telomere length (TL) is another genetic-related biomarker (Fig. 4.2C). Telomeres are repetitive DNA sequences placed on the ends of chromosomes that protect the end of the chromosome from deterioration or fusion with nearby chromosomes. Telomeres are thought to be a marker of cellular aging. Within human tissues telomeres shorten with each cell division and as chronological age advances (Eisenberg, 2011). There are mixed results for the comparisons between TL and biological age or mortality (Mather, Jorm, Parslow, & Christensen, 2011) as well as inconsistent results related to the predictors of TL and telomere shortening such as stress and disadvantage (Oliveira et al., 2016). Despite a great deal of interest and initial promise in TL as a biomarker, technological issues related to collection and measurement have complicated the scientific landscape and our ability to both discover and replicate TL associations with age-related exposures or outcomes. The main challenge in telomere research has been the development of accurate and reliable measurement methods to achieve sensitive and reliable TL measurement across labs (Mensà et al., 2019). Addressing these issues would help avoid major biases in association studies involving TL and a number of outcomes, especially those focusing on psychological and biobehavioral variables (Montpetit et al., 2014). In addition, researchers need to publish the full details of their methods and the quality control procedures they use—inclusion of lab replicates, batch effects, laboratory reagents, and DNA concentration—so that potential errors generated by confounding variables are explicit.

Gene by environment interactions

How the environment interacts with one's genes to realize an outcome, or conversely, how genes help to select us into different environments is of primary interest to social scientists. Gene−environment interaction (GxE) research—studying how genetic predispositions interact with environmental factors to contribute to complex disease and behavioral outcomes—holds great promise and has become an important area of study across multiple disciplines. However, challenges such as statistical concerns with modeling GxE (e.g., high false positive rates), gene−environment correlation, low power, and publication bias plague some of the research in this area, particularly the candidate gene−environment interaction literature (Dick et al., 2015).

An accumulating body of research has demonstrated that the importance of genetic influences can vary dramatically as a function of environmental context; importantly and alternatively phrased, the importance of environmental influences can vary dramatically as a function of genetic factors. For example, GxE studies have examined how state and peer effects modify the genetic predisposition for substance use (Boardman, 2009; Do & Maes, 2016), how genetic factors interact with childhood socioeconomic status in determining educational outcomes (Papageorge & Thom, 2018), and how genes and environments together shape how we age cognitively (Reynolds, 2014). This ever-growing body of GxE research grew initially from the field of twin research. Not until the Science publication by Caspi et al. (2003) that examined genetic sensitivity to stressful life events (Caspi et al., 2003), specifically variations in a DNA sequence [a polymorphism in the serotonin-transporter-linked polymorphic region (5-HTTLPR)], did GxE become more widespread outside of the field of behavioral genetics.

Population-based studies add to our understanding of GxE by showing how factors across the life course and genetics jointly contribute to later life outcomes (Domingue et al., 2017). Examples include work showing that gene−stress interactions may influence the aging brain and contribute to neuropsychiatric phenotypes in later life. Multiple gene−stress interactions may act in tandem, along with other environmental factors and aging-related brain processes, to induce changes in gene expression patterns across brain regions that play critical roles in the regulation of mood and cognition, and in the development of neuropsychiatric syndromes (Zannas, McQuoid, Steffens, Chrousos, & Taylor, 2012). However, few other studies have examined GxE interactions in older individuals, let alone in the population-based and population-representative studies outlined in Table 4.1.

It is important to note that GxE can be examined at several different genetic levels. Using methodologies such as family, twin, and adoption studies, GxE can examine "latent" genetic influences in which the importance of genetic factors is estimated statistically by phenotypic similarity across individuals with different degrees of genetic and environmental sharing. Similarly, one could examine GxE using polygenic or other genome-wide scores. Genome-wide methods examine the interaction of the environment with genetic effects across the entire genome on a phenotype, that is, the total contribution of all genes influencing the phenotype. This method, however, does not allow for heterogeneity in the direction of the effect by gene or gene region, nor does this method pinpoint biological mechanisms due to the aggregation of effects and so might be limited in how biologically informative it can be. In contrast, preliminary GxE research in fields outside of behavioral genetics have targeted specific candidate genes. These studies test whether the association of a genetic variant identified *a priori* with a given outcome varies across different environments. Statistical concerns and issues with reproducibility have moved the field away from the candidate GxE approach toward others that take into account the dynamism of GxG interaction as well as GxE interaction. Investigating the interplay of genes and the environment in human behavior and disease will require more advanced statistical and computationally efficient methods. More interdisciplinary work integrating molecular biology, environmental sciences, bioinformatics, and statistical and computational methods is likely to advance research in this area in the years to come.

Limitations

It is clear that the recent massive expansion of genetic research is generating a large literature that has been transformative in how we think about the influences on health and behavior. While the field is usually careful to address many important limitations such as multiple testing and population stratification, many other limitations, several quite familiar to the world of social science, have not yet received as much attention as they should. Most genome-wide studies to date, as well as a tremendous amount of the work and methods derived from GWAS, have largely skewed samples in terms of race/ancestry but also by SES (Mills & Rahal, 2019; Popejoy & Fullerton, 2016). While this is slowly changing, the vast majority of work uses higher SES samples from European ancestry populations. This is particularly problematic because allele frequencies differ by ancestry (population stratification) and thus findings from GWASs that associate certain alleles to specific outcomes do not translate across ancestral groups. Similarly, the creation of summary scores that aggregate the effect of genetic associations across the genome, like PGSs, using GWAS of European samples will not have the same predictive power across other ancestry groups complicating the ability to do race/ethnicity-based health disparities work. Moreover, this same problem exists for work in admixture populations combining two or more ancestral groups. Because much more genetic variation exists in non-European populations, we will have to invest in creating much larger samples to identify the same number of significant genome-wide associations. While efforts are underway to increase the representativeness of available genetic data in terms of race/ethnicity, not enough attention is given to the lack of representativeness along other dimensions such as education, income, and wealth. It is important to note that the strength of genetic associations with outcomes of interest is also driven by the context in which they are observed. While several population-based studies have genetic data (e.g., Fragile Families and Child Wellbeing Study; the National Study of Adolescent Health; the Health and Retirement Study; the English Longitudinal Study of Ageing; the Wisconsin Longitudinal Study), the composite samples of most of the genetic consortia are not population representative. This situation is only magnified with the introduction of large samples like the UK Biobank which are significantly skewed and often contribute a substantial proportion to the GWAS meta-analyses (Fry et al., 2017).

Differential consent and selective mortality also play roles in the genomic sciences in much the same way they do in other social and health sciences. Differential selection by consent into a genomic dataset can influence results from genomic analyses. Selective nonparticipation can be related to the same traits, behaviors, and environmental contexts that we most want to focus on in our genomic research. In addition, mortality selection—that genomic data collected at a particular age and point in time represents the subset of birth cohort members who have survived to the time of data collection—can also lead to bias, especially when the trait of interest is highly associated with mortality (Domingue et al., 2017). Selection can have effects not only on studies of individual variants (e.g., GWAS), but also on gene−environment interaction studies and downstream use of GWAS data such as PGS. In general, researchers should consider the association of their outcome of interest with inclusion in the genetic sample when making claims about the generalizability of their work.

Limitations in genomic research also come from how specifically phenotypes are defined in discovery GWAS. A sample size with sufficient statistical power is crucial to genetic association studies in order to detect causal genes of human complex diseases and behavioral traits. Genome-wide association studies require very large sample sizes to achieve adequate statistical power (Hong & Park, 2012; Visscher et al., 2017). It can be difficult to amass sufficiently large discovery and replication samples for GWAS with consistently measured phenotypes. As a result of the phenotype harmonization, also known as minimal phenotyping, that is often required to pool large enough samples, the specificity of the phenotype can often be lost and can reduce our ability to identify genetic variants or pathways associated with the outcomes of interest.

Ethical issues

A chapter on genomics in social science research must also address the ethical issues that surround this work. Most of the ethical arguments surround the use of genomics to investigate whether genetic differences by ancestral group explain observed differences in health, behaviors, and other outcomes between self-identified race/ethnic groups. Researchers are interested in knowing whether genetics matter more than the environmental influences that we also know matter. On one side of the argument researchers claim that the use of genomics in social science and health disparities research "naturalizes" racial and ethnic differences and disincentivizes researchers from investigation into the complex ways in which social inequality and experiences, such as stress and discrimination, interact with biology to influence outcomes (Braun, 2002; Cooper, Kaufman, & Ward, 2003). At the root of this concern is the widespread, but outdated, idea that genetics is destiny—that genes can accurately predict complex behaviors and outcomes regardless of their environment. Some researchers and funders have leaned away from funding work in this area fearing that it will lead to a new age of eugenics, or increased discrimination against already marginalized groups (Hayden, 2013). It does not help that this genetic determinism is fueled by a consumer and research environment that is promising individualized health risk assessments and personalized medicine before there is sufficient proof to deliver [in addition to genetically inspired travel itineraries and music playlists (Mahdawi, 2019)]. On the other side of the argument, scientists reinforce that race is a social construct, highly correlated with ancestry, which cannot be used as a biological classification. Moreover, even if genes can predict racial differences in outcomes like IQ or disease, they can do so because genes are good predictors of social context (Conley & Fletcher, 2017).

In order to address these issues, researchers must acknowledge these ethical issues related to genomics and take part in the broader conversation about their work. If the potential exists to misuse the science we produce we should not avoid engaging in scientific inquiry; rather we need to create principles to guide the use of race/ethnicity and ancestry in genetic research and continue an open, public dialogue that ensures responsible use (Hayden, 2013; Lee et al., 2008). Censoring science or ignoring its implications ultimately only serves to lend credit to the notion that there is something biologically deterministic about race, when in fact the ability to control for genotype actually highlights the effect of environmental and social processes like discrimination. If biological and genetic differences between populations can be controlled for in our models, then we can see more clearly the influence of environmental (nongenetic) processes such as structural racism (Conley & Fletcher, 2017). It is here, at the intersection between genomics and race, where the expertise of social science can lend the most to the field of genomics.

On the horizon—mitochondrial DNA and gene expression

The field of genetics moves quickly and new measures and biomarkers develop as we better understand how to interpret and analyze the volumes of DNA data we collect. Social science studies are adding genomic data as they become cheaper and easier to produce on a large scale. Measures beyond genotype and methylation data that show promise include mitochondrial DNA, gene expression data, and data from the transcriptome. Together these measures may help us to identify better aging trajectories and possibly life-extending strategies (Tarkhov et al., 2019).

Mitochondria are the main sources of cellular ATP, an important energy source (Kujoth et al., 2005). Mitochondria contain their own DNA (mtDNA). Mutations and alterations to mtDNA may play a role in aging. Older people have higher oxidative damage to mtDNA. mtDNA abundance decreases with age and correlates with lower content of mRNA transcripts that encode mitochondrial protein (Short et al., 2005). There are significant differences in the number of mtDNA deletions between old and young people (Kraytsberg et al., 2006; Mohamed et al., 2006). Some distinct patterns of mtDNA have been found to associate with longevity (Bilal et al., 2008; Castri et al., 2009; De Benedictis et al., 1999; Niemi et al., 2005; Zhang et al., 2003).

Genes encode proteins and proteins dictate cell function. The genes expressed in a particular cell determine what that cell can do. Each step in the flow of information from DNA to RNA to protein provides the cell with a potential control point for self-regulating its functions. Protein production starts at transcription (DNA to RNA) and continues with translation (RNA to protein), and control of these processes plays a critical role in determining what proteins are present in a cell and in what amounts (O'Connor & Adams, 2010). Gene expression is also an important source of evolutionary change and abnormal gene expression has been implicated in the pathogenesis of numerous diseases (Knight, 2005; Storey et al., 2007; Yan & Zhou, 2004). Studying the transcriptome (RNA expressed from the genome) is one way to examine gene expression. RNA-Seq (RNA sequencing), also called whole transcriptome shotgun sequencing (*WTSS*), uses next-generation sequencing to measure the presence and quantity of RNA at a given point in time, a marker of the ever-changing transcriptome (Chu & Corey, 2012; Wang, Gerstein, & Snyder, 2009) Increasingly, transcriptome analysis has become central to functionally correlate genetic variations identified in GWAS to disease phenotypes such as cancer and neurodegenerative diseases (Costa, Aprile, Esposito, & Ciccodicola, 2013). In addition, gene expression and gene regulation are interesting potential targets for research on the ways in which social and environmental influences can affect our biology. As we learn more about how gene expression variation is apportioned within populations and population subgroups, we will learn more about the ways in which genes interact with and respond to environments through expression.

We expect modifiable genetics, such as DNA methylation and gene expression, to be influenced more by the environment than the genomic structure (DNA) of individuals. As we come to understand the complex interplay between genomics and the environment, it becomes increasingly important to understand how to characterize the social and environmental context. While genetics has made important contributions to behavioral and social science research, social science research can lend careful sociocontextual measurement methods and theory to genomic-based research on aging. Aging represents the accumulation of a lifetime of experiences and exposures, both social and biological, and only with an integrated approach will we begin to truly understand the processes that underlie healthy aging across the life course.

References

Bell, J. T., & Spector, T. D. (2011). A twin approach to unraveling epigenetics. *Trends in Genetics: TIG*, *27*(3), 116–125. Available from https://doi.org/10.1016/j.tig.2010.12.005.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57* (1), 289–300.

Beyene, J., Tritchler, D., Asimit, J. L., & Hamid, J. S. (2009). Gene- or region-based analysis of genome-wide association studies. *Genetic Epidemiology*, *33*(Suppl. 1), S105–110. Available from https://doi.org/10.1002/gepi.20481.

Bilal, E., Rabadan, R., Alexe, G., Fuku, N., Ueno, H., Nishigaki, Y., et al. (2008). Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan. *Plos One*, *3*(6). Available from https://doi.org/10.1371/journal.pone.0002421, ARTN e2421.

Boardman, J. D. (2009). State-level moderation of genetic tendencies to smoke. *American Journal of Public Health*, *99*(3), 480–486. Available from https://doi.org/10.2105/AJPH.2008.134932.

Braun, L. (2002). Race, ethnicity, and health: Can genetics explain disparities? *Perspectives in Biology and Medicine*, *45*(2), 159–174.

Breitling, L. P., Saum, K.-U., Perna, L., Schöttker, B., Holleczek, B., & Brenner, H. J. (2016). Frailty is associated with the epigenetic clock but not with telomere length in a German cohort. *Clinical Epigenetics*, *8*(1), 21.

Buil, A., Martinez-Perez, A., Perera-Lluna, A., Rib, L., Caminal, P., & Soria, J. M. (2009). A new gene-based association test for genome-wide association studies. *BMC Proceedings*, *3*(Suppl. 7), S130.

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: Moderation by a polymorphism in the 5-HTT gene. *Science (New York, N.Y.)*, *301*(5631), 386–389. Available from https://doi.org/10.1126/science.1083968.

Castri, L., Melendez-Obando, M., Villegas-Palma, R., Barrantes, R., Raventos, H., Pereira, R., et al. (2009). Mitochondrial polymorphisms are associated both with increased and decreased longevity. *Human Heredity*, *67*(3), 147–153. Available from https://doi.org/10.1159/000181152.

Chen, X., Wang, L., Smith, J. D., & Zhang, B. (2008). Supervised principal component analysis for gene set enrichment of microarray data with continuous or survival outcomes. *Bioinformatics*, *24*(21), 2474–2481. Available from https://doi.org/10.1093/bioinformatics/btn458.

Chu, Y., & Corey, D. R. (2012). RNA sequencing: Platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*, *22*(4), 271–274. Available from https://doi.org/10.1089/nat.2012.0367.

Cirillo, E., Parnell, L. D., & Evelo, C. T. (2017). A review of pathway-based analysis tools that visualize genetic variants. *Frontiers in Genetics*, *8*, 174. Available from https://doi.org/10.3389/fgene.2017.00174.

Clinical and translational science (2nd ed.) (2016). Amsterdam.

Conley, D., & Fletcher, J. (2017). What both the left and right get wrong about race. *Nautilus*. Available from http://nautil.us/issue/48/chaos/what-both-the-left-and-right-getwrong-about-race (accessed on 22 May 2020).

Cooper, R. S., Kaufman, J. S., & Ward, R. (2003). Race and genomics. *New England Journal of Medicine*, *348*(12), 1166–1170. Available from https://doi.org/10.1056/NEJMsb022863.

Costa, V., Aprile, M., Esposito, R., & Ciccodicola, A. (2013). RNA-Seq and human complex diseases: Recent accomplishments and future perspectives. *European Journal of Human Genetics*, *21*(2), 134–142. Available from https://doi.org/10.1038/ejhg.2012.129.

Cunliffe, V. T. (2016). The epigenetic impacts of social stress: How does social adversity become biologically embedded? *Epigenomics*, *8*(12), 1653–1669. Available from https://doi.org/10.2217/epi-2016-0075.

De Benedictis, G., Rose, G., Carrieri, G., De Luca, M., Falcone, E., Passarino, G., et al. (1999). Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. *Faseb Journal*, *13*(12), 1532–1536.

Dick, D. M., Agrawal, A., Keller, M. C., Adkins, A., Aliev, F., Monroe, S., et al. (2015). Candidate gene-environment interaction research: Reflections and recommendations. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *10*(1), 37–59. Available from https://doi.org/10.1177/1745691614556682.

Do, E., & Maes, H. (2016). Narrative review of genes, environment, and cigarettes. *Annals of Medicine*, *48*(5), 337–351. Available from https://doi.org/10.1080/07853890.2016.1177196.

Domingue, B. W., Belsky, D. W., Harrati, A., Conley, D., Weir, D. R., & Boardman, J. D. (2017). Mortality selection in a genetic sample and implications for association studies. *International Journal of Epidemiology*, *46*(4), 1285–1294. Available from https://doi.org/10.1093/ije/dyx041.

Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, *11*(1), 587. Available from https://doi.org/10.1186/1471-2105-11-587.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *Plos Genetics*, *9*(3). Available from https://doi.org/10.1371/journal.pgen.1003348, ARTN e1003348.

Dupont, C., Armant, D. R., & Brenner, C. A. (2009). Epigenetics: Definition, mechanisms and clinical perspective. *Seminars in Reproductive Medicine*, *27*(5), 351–357. Available from https://doi.org/10.1055/s-0029-1237423.

Eisenberg, D. T. A. (2011). An evolutionary review of human telomere biology: The thrifty telomere hypothesis and notes on potential adaptive paternal effects. *American Journal of Human Biology*, *23*(2), 149–167. Available from https://doi.org/10.1002/ajhb.21127.

Flanagan, J. M. (2015). Epigenome-wide association studies (EWAS): Past, present, and future. *Methods in Molecular Biology (Clifton, N.J.)*, *1238*, 51–63. Available from https://doi.org/10.1007/978-1-4939-1804-1_3.

Field, A. E., Robertson, N. A., Wang, T., Havas, A., Ideker, T., & Adams, P. D. (2018). DNA methylation clocks in aging: categories, causes, and consequences. *Molecular Cell*, *71(6)*, 882–895. Available from https://doi.org/10.1016/j.molcel.2018.08.008.

Fraga, M. F., Agrelo, R., & Esteller, M. (2007). Cross-talk between aging and cancer: The epigenetic language. *Annals of the New York Academy of Sciences*, *1100*, 60–74. Available from https://doi.org/10.1196/annals.1395.005.

Fraga, M. F., & Esteller, M. (2007). Epigenetics and aging: The targets and the marks. *Trends in Genetics: TIG*, *23*(8), 413–418. Available from https://doi.org/10.1016/j.tig.2007.05.008.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., et al. (2017). Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology*, *186*(9), 1026–1034. Available from https://doi.org/10.1093/aje/kwx246.

Gale, C. R., Marioni, R. E., Harris, S. E., Starr, J. M., & Deary, I. J. (2018). DNA methylation and the epigenetic clock in relation to physical frailty in older people: The Lothian Birth Cohort 1936. *Clinical Epigenetics*, *10*(1), 101. Available from https://doi.org/10.1186/s13148-018-0538-4.

Gao, Q., He, Y., Yuan, Z., Zhao, J., Zhang, B., & Xue, F. (2011). Gene- or region-based association study via kernel principal component analysis. *BMC Genetics*, *12*, 75. Available from https://doi.org/10.1186/1471-2156-12-75.

Gauderman, W. J., Murcray, C., Gilliland, F., & Conti, D. V. (2007). Testing association between disease and multiple SNPs in a candidate gene. *Genetic Epidemiology*, *31*(5), 383–395. Available from https://doi.org/10.1002/gepi.20219.

Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467* (7319), 1061–1073. Available from https://doi.org/10.1038/nature09534.

Han, F., & Pan, W. (2010). A data-adaptive sum test for disease association with multiple common or rare variants. *Human Heredity*, *70*(1), 42–54. Available from https://doi.org/10.1159/000288704.

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., et al. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell*, *49*(2), 359–367. Available from https://doi.org/10.1016/j.molcel.2012.10.016.

Hayden, E. C. (2013). Ethics: Taboo genetics. *Nature News*, *502*(7469), 26. Available from https://doi.org/10.1038/502026a.

Herskind, A. M., McGue, M., Holm, N. V., Sörensen, T. I., Harvald, B., & Vaupel, J. W. (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human Genetics*, *97*(3), 319–323.

Hong, E. P., & Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics*, *10*(2), 117–122. Available from https://doi.org/10.5808/GI.2012.10.2.117.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10). Available from https://doi.org/10.1186/gb-2013-14-10-r115, ARTN R115.

Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., et al. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging*, *10*(7), 1758.

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, *19*(6), 371. Available from https://doi.org/10.1038/s41576-018-0004-3.

Horvath, S., & Ritz, B. R. (2015). Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging-Us*, *7*(12), 1130–1142. Available from https://doi.org/10.18632/aging.100859.

Ionita-Laza, I., Buxbaum, J. D., Laird, N. M., & Lange, C. (2011). A new testing strategy to identify rare variants with either risk or protective effect on disease. *Plos Genetics*, *7*(2), e1001289. Available from https://doi.org/10.1371/journal.pgen.1001289.

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., & Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. *American Journal of Human Genetics*. Available from https://doi.org/10.1016/j.ajhg.2013.04.015.

Jabbari, K., & Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene*, *333*, 143–149. Available from https://doi.org/10.1016/j.gene.2004.02.043.

Jaffe, A. E., & Irizarry, R. A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome*

Biology, *15*(2), R31. Available from https://doi.org/10.1186/gb-2014-15-2-r31.

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, *51*(3), 404−413. Available from https://doi.org/10.1038/s41588-018-0311-9.

Khera, A. V., Emdin, C. A., Drake, I., Natarajan, P., Bick, A. G., Cook, N. R., et al. (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. *The New England Journal of Medicine*, *375*(24), 2349−2358. Available from https://doi.org/10.1056/NEJMoa1605086.

Knight, J. C. (2005). Regulatory polymorphisms underlying complex disease traits. *Journal of Molecular Medicine*, *83*(2), 97−109. Available from https://doi.org/10.1007/s00109-004-0603-7.

Kraytsberg, Y., Kudryavtseva, E., McKee, A. C., Geula, C., Kowall, N. W., & Khrapko, gK. J. N. (2006). Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nature Genetics*, *38*(5), 518.

Kujoth, G. C., Hiona, A., Pugh, T. D., Someya, S., Panzer, K., Wohlgemuth, S. E., et al. (2005). Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science*, *309* (5733), 481−484. Available from https://doi.org/10.1126/science.1112125.

Lamb, N. (2007). *Epigenetics*. Hudson Alpha Institute for Biotechnology.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112. Available from https://doi.org/10.1038/s41588-018-0147-3.

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., et al. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American Journal of Human Genetics*, *91*(2), 224−237. Available from https://doi.org/10.1016/j.ajhg.2012.06.007.

Lee, S., Wu, M. C., & Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, *13*(4), 762−775. Available from https://doi.org/10.1093/biostatistics/kxs014.

Lee, S. S.-J., Mountain, J., Koenig, B., Altman, R., Brown, M., Camarillo, A., et al. (2008). The ethics of characterizing difference: Guiding principles on using racial categories in human genetics. *Genome Biology*, *9*(7), 404. Available from https://doi.org/10.1186/gb-2008-9-7-404.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., et al. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging-Us*, *10*(4), 573−591. Available from https://doi.org/10.18632/aging.101414.

Madsen, B. E., & Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *Plos Genetics*, *5* (2), e1000384. Available from https://doi.org/10.1371/journal.pgen.1000384.

Mahdawi, A. (June 04, 2019). DNA-based holidays encourage a dangerous flirtation with race. *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2019/jun/04/dna-based-holidays-encourage-a-dangerous-flirtation-with-race

Marchini, J., & Band, G. (2019). SNPTEST v2.5.4-beta3.

Marioni, R. E., Shah, S., McRae, A. F., Chen, B. H., Colicino, E., Harris, S. E., et al. (2015). DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology*, *16*. Available from https://doi.org/10.1186/s13059-015-0584-6, ARTN 25.

Mather, K. A., Jorm, A. F., Parslow, R. A., & Christensen, H. (2011). Is telomere length a biomarker of aging? A review. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*, *66*(2), 202−213. Available from https://doi.org/10.1093/gerona/glq180.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279−1283. Available from https://doi.org/10.1038/ng.3643.

Mensà, E., Latini, S., Ramini, D., Storci, G., Bonafè, M., & Olivieri, F. (2019). The telomere world and aging: Analytical challenges and future perspectives. *Ageing Research Reviews*, *50*, 27−42. Available from https://doi.org/10.1016/j.arr.2019.01.004.

Mills, M. C., & Rahal, C. (2019). A scientometric review of genome-wide association studies. *Communications Biology*, *2*(1), 9. Available from https://doi.org/10.1038/s42003-018-0261-x.

Mohamed, S. A., Hanke, T., Erasmi, A. W., Bechtel, M. J. F., Scharfschwerdt, M., Meissner, C., et al. (2006). Mitochondrial DNA deletions and the aging heart. *Experimental Gerontology*, *41* (5), 508−517. Available from https://doi.org/10.1016/j.exger.2006.03.014.

Montpetit, A. J., Alhareeri, A. A., Montpetit, M., Starkweather, A. R., Elmore, L. W., Filler, K., et al. (2014). Telomere length: A review of methods for measurement. *Nursing Research*, *63*(4), 289−299. Available from https://doi.org/10.1097/NNR.0000000000000037.

National Heart, Lung and Blood Institute. (2018). Trans-Omics for Precision Medicine (TOPMed) Program. Available from https://www.nhlbiwgs.org/

National Human Genome Research Institute. (2015). Genome-Wide Association Studies Fact Sheet. Available from https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet.

National Human Genome Research Institute. (2018). Genetics vs. Genomics Fact Sheet Genome.gov.

Naumova, O. Y., Rychkov, S. Y., Kornilov, S. A., Odintsova, V. V., Anikina, V. O., Solodunova, M. Y., et al. (2019). Effects of early social deprivation on epigenetic statuses and adaptive behavior of young children: A study based on a cohort of institutionalized infants and toddlers. *Plos One*, *14*(3), e0214285. Available from https://doi.org/10.1371/journal.pone.0214285.

Neale, B. M., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., et al. (2011). Testing for an unusual distribution of rare variants. *Plos Genetics*, *7*(3), e1001322. Available from https://doi.org/10.1371/journal.pgen.1001322.

Niemi, A.-K., Moilanen, J. S., Tanaka, M., Hervonen, A., Hurme, M., Lehtimäki, T., et al. (2005). A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. *European Journal of Human Genetics*, *13*(2), 166.

Notterman, D. A., & Mitchell, C. (2015). Epigenetics and understanding the impact of social determinants of health. *Pediatric Clinics of North America*, *62*(5), 1227−1240. Available from https://doi.org/10.1016/j.pcl.2015.05.012.

O'Connor, C. M., & Adams, J. U. (2010). *Essentials of cell biology*. Cambridge, MA: NPG Education.

Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Mark, A. F., et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6).

Oliveira, B. S., Zunzunegui, M. V., Quinlan, J., Fahmi, H., Tu, M. T., & Guerra, R. O. (2016). Systematic review of the association between chronic social stress and telomere length: A life course perspective. *Ageing Research Reviews*, *26*, 37−52. Available from https://doi.org/10.1016/j.arr.2015.12.006.

Papageorge, N.W., & Thom, K. (2018). *Genes, education, and labor market outcomes: Evidence from the Health and Retirement Study* (25114). Retrieved from <http://www.nber.org/papers/w25114>.

Perna, L., Zhang, Y., Mons, U., Holleczek, B., Saum, K. U., & Brenner, H. (2016). Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clinical Epigenetics*, *8*. Available from https://doi.org/10.1186/s13148-016-0228-z, ARTN 64.

Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H.-J., Wood, A. R., Yang, J., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications*, *6*, 5890. Available from https://doi.org/10.1038/ncomms6890.

Phillips, N. L. H., & Roth, T. L. (2019). Animal models and their contribution to our understanding of the relationship between environments, epigenetic modifications, and behavior. *Genes (Basel)*, *10* (1). Available from https://doi.org/10.3390/genes10010047.

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature News*, *538*(7624), 161. Available from https://doi.org/10.1038/538161a.

Price, A. L., Kryukov, G. V., de Bakker, P. I., Purcell, S. M., Staples, J., Wei, L. J., & Sunyaev, S. R. (2010). Pooled association tests for rare variants in exon-resequencing studies. *American Journal of Human Genetics*, *86*(6), 832–838. Available from https://doi.org/10.1016/j.ajhg.2010.04.005.

Qiao, B., Huang, C. H., Cong, L., Xie, J., Lo, S. H., & Zheng, T. (2009). Genome-wide gene-based analysis of rheumatoid arthritis-associated interaction with PTPN22 and HLA-DRB1. *BMC Proceedings*, *3*(Suppl. 7), S132.

Reynolds, C.A. (2014). How genes and environments (co)act to shape how we age cognitively. <https://www.apa.org>.

Rodríguez-Rodero, S., Fernández-Morera, J. L., Menéndez-Torre, E., Calvanese, V., Fernández, A. F., & Fraga, M. F. (2011). Aging genetics and aging. *Aging and Disease*, *2*(3), 186–195.

Short, K. R., Bigelow, M. L., Kahl, J., Singh, R., Coenen-Schimke, J., Raghavakaimal, S., & Nair, K. S. (2005). Decline in skeletal muscle mitochondrial function with aging in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5618–5623. Available from https://doi.org/10.1073/pnas.0501559102.

Storey, J. D., Madeoy, J., Strout, J. L., Wurfel, M., Ronald, J., & Akey, J. M. (2007). Gene-expression variation within and among human populations. *American Journal of Human Genetics*, *80*(3), 502–509.

Tarkhov, A. E., Alla, R., Ayyadevara, S., Pyatnitskiy, M., Menshikov, L. I., Shmookler Reis, R. J., & Fedichev, P. O. (2019). A universal transcriptomic signature of age reveals the temporal scaling of *Caenorhabditis elegans* aging trajectories. *Science Reports*, *9*(1), 7368. Available from https://doi.org/10.1038/s41598-019-43075-z.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., et al. (2010). Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Research*, *20*(4), 440–446. Available from https://doi.org/10.1101/gr.103606.109.

Tzeng, J. Y., Zhang, D., Pongpanich, M., Smith, C., McCarthy, M. I., Sale, M. M., et al. (2011). Studying gene and gene-environment effects of uncommon and common variants on continuous traits: A marker-set approach using gene-trait similarity regression. *American Journal of Human Genetics*, *89*(2), 277–288. Available from https://doi.org/10.1016/j.ajhg.2011.07.007.

Vaiserman, A. (2018). Developmental tuning of epigenetic clock. *Frontiers in Genetics*, *9*. Available from https://doi.org/10.3389/fgene.2018.00584.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature Reviews. Genetics*, *9*(4), 255–266. Available from https://doi.org/10.1038/nrg2322.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS discovery: Biology, function, and translation. *The American Journal of Human Genetics*, *101*(1), 5–22. Available from https://doi.org/10.1016/j.ajhg.2017.06.005.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. Available from https://doi.org/10.1038/nrg2484.

Webster, A. P., Plant, D., Ecker, S., Zufferey, F., Bell, J. T., Feber, A., et al. (2018). Increased DNA methylation variability in rheumatoid arthritis-discordant monozygotic twins. *Genome Medicine*, *10*(1), 64. Available from https://doi.org/10.1186/s13073-018-0575-9.

Willcox, B. J., Willcox, D. C., He, Q. M., Curb, J. D., & Suzuki, M. (2006). Siblings of Okinawan centenarians share lifelong mortality advantages. *Journals of Gerontology Series a-Biological Sciences and Medical Sciences*, *61*(4), 345–354. Available from https://doi.org/10.1093/gerona/61.4.345.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, *89*(1), 82–93. Available from https://doi.org/10.1016/j.ajhg.2011.05.029.

Yan, H., & Zhou, W. (2004). Allelic variations in gene expression. *Current Opinion in Oncology*, *16*(1), 39–43. Available from https://doi.org/10.1097/00001622-200401000-00008.

Zannas, A. S., McQuoid, D. R., Steffens, D. C., Chrousos, G. P., & Taylor, W. D. (2012). Stressful life events, perceived stress, and 12-month course of geriatric depression: Direct effects and moderation by the 5-HTTLPR and COMT Val158Met polymorphisms. *Stress-the International Journal on the Biology of Stress*, *15*(4), 425–434. Available from https://doi.org/10.3109/10253890.2011.634263.

Zhang, J., Asin-Cayuela, J., Fish, J., Michikawa, Y., Bonafé, M., Olivieri, F., et al. (2003). Strikingly higher frequency in centenarians and twins of mtDNA mutation causing remodeling of replication origin in leukocytes. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(3), 1116–1121.

Zheng, Y., Joyce, B. T., Colicino, E., Liu, L., Zhang, W., Dai, Q., et al. (2016). Blood epigenetic age may predict cancer incidence and mortality. *Cancer Research*, *76*. Available from https://doi.org/10.1158/1538-7445.Am2016-4480.