4-2020

The Era of Commercialized Genetics: Examining the Intersection of DNA, Identity, and Personal Origin

Joanna Hutchins 20
DePauw University

The Era of Commercialized Genetics
Examining the Intersection of DNA, Identity, and Personal Origin

Joanna Hutchins
DePauw University
Honor Scholar Program
Class of 2020

Sponsor
Dr. Joseph Heithaus

Committee
Dr. Dana Dudle
Dr. Mahdis Azarmandi

ACKNOWLEDGEMENTS

I would like to sincerely thank *Professor Joseph Heithaus* for investing endless time and thought in guiding me through this project. Amidst the unknowns of our current world state, our meetings provided both structure and release. He may joke about why a poet served as the sponsor for a thesis on genetics, but it is precisely this perspective that enriched our conversations about the meaning of personhood and the social life of science.

I would also like to thank *Professor Dana Dudle* for talking me through the conception of this project, for helping me navigate the intricate concepts of evolutionary biology, and for vitalizing conversations about everything I was learning.

I want to extend my deep gratitude to *Professor Mahdis Azarmandi*, who showed me the vital importance of the social sciences, challenged me to identify and reframe my own embedded ideas and perspectives, and helped me see the value in my own self and work.

Throughout my time at DePauw, I have been fascinated by the interconnectedness of my classes in science, art, history, and humanities, how subjects so seemingly dissimilar continued to interdependently inform my perspective and augment my approach within each. And while I progressively honed in on my major in Kinesiology throughout my four years, I felt apprehensive that I was not fully tapping the unique experiences of the liberal arts. Friends spoke of incredible professors, those like Prof. Heithaus, Prof. Dudle, and Prof. Azarmandi, and unforgettable learning experiences across departments that I did not have the chance to intersect. Perhaps this thesis was my way of bringing these departments, and immensely more so these professors, to me, because the most meaningful part of my DePauw education will always be relationships.

Preface

When I first set out to write this thesis, I hoped to translate the complexity of modern genomics into a couple telling sentences that could describe across readerships exactly what genetics can and cannot conclusively tell us about ancestry.

As I delved into the history of scientific racism, trying to decipher lines of continuity between past and present rhetoric used in primary literature, I contemplated the limitations of my perspective. Making constant decisions, whether large or small, on which topics to cover and how to orient my message, I thought about how my identity as a white woman informed those decisions. I wondered how my decisions would have changed, which histories I would have highlighted.

There is a complex relationship between the idea of genes and personal origin. Both have ties to inheritance, the passing down of genetic code, cultural tradition or family history. Our sense of origin, and its relationship to identity, lives within a social world. It is formed and informed by experience, how we navigate the world. Genes lack intrinsic social meaning but, in many cases, have been endowed social authority.

For me, personal origin arrives with a bit of ambiguity. It seems logical that answering the question of "who am I" necessarily follows "where did I come from?" But how long ago is still relevant to my everyday realities? Growing up, I learned to view my heritage is an assemblage of German, Venezuelan, English, Portuguese, Trinidadian - each story conceptualized as a nationality. I don't feel I can make full claim to any of these as component to my identity. I am not bilingual like my mom and her mom and her mom. I never experienced the reformation of national identity through immigration, like my mother. And at the same time, I do feel that my upbringing was impacted by the cultural differences between my mother and my father. I've always held my mom's Trinidadian origins with a sense of pride and a sort of distant closeness. But I have ultimately navigated this world as a white American in a hegemonic system.

The conversation is one of the most complex concepts I have ever engaged with. It is a gateway into discussions of race and racism throughout history, of science as imbued with authority, of personal origin tied to culture versus genealogy versus DNA, of critical examinations centering semantics and discourse, of continuity in seemingly cyclical reproductions of the past in the present, of how ancestry is imagined and reimagined in the era of commercialized genetics. While I cannot feasibly capture them all, giving each the nuanced attentiveness and breadth of research they require, I hope that my conversations throughout this thesis lead to more and highlight the inter-connectivity of ideas spanning the humanities, history, social sciences, memory studies, etc. that can be found in any research question, even one prodding at the assumed detached objectivity of natural science.

TABLE OF CONTENTS

INTRODUCTION

When Elizabeth Warren Said…

When Elizabeth Warren released her personal DNA test report to corroborate family folklore of Native American ancestry, she landed herself at the intersection of a complex cultural, sociological, and biological debate that exposed a deep, conceptual fissure based in false, or at least oversimplified, understandings of race, nation, and culture.

This decision positioned genetic testing on par with cultural identity, implying that identity is "discoverable" via test reports, rather than crafted through community, embedded in experience, and sustained by collective memory. Slurs espoused by Donald Trump, who called her "Pocahontas" in a slipshod attempt of defamation, further incited tensions surrounding Warren's ancestral claims (Astor, 2018). Her decision to get genetic testing led to the revelation that she had a Native American ancestor likely six generations ago. But Warren's misconceived intention to embrace an imagined ancestry that she believed might be confirmed by science led to a damaging double bind. If her priority was to placate the President's taunts, she neglected the negative connotation and caricature that critically underly white conceptions of Indigenous peoples. If her purpose was to prove her family's stories and claims to Native American ancestry, then she undermined what Native American, and more specifically Cherokee, identity means, delegitimizing tribal citizenship. Cherokee Nation citizenship is founded upon established laws and culture as well as genealogy. DNA test results are insubstantial according to this understanding, reflecting neither the formality of law nor the reality of human experience.

Cultural identity is not inherently tied to genetic ancestry, as the former is an active embodiment and a way of life while the latter traces vestiges of biological data in an attempt to systematize a social construct. Genetic testing has a crucial historical dimension, and although increasingly advanced technology has shifted its mechanisms and reframed the conversation , the rhetoric surrounding genetic ancestry is reminiscent of outmoded taxonomies that were leveraged to justify racism. Typically, the strong association of blood relations being the sole determinant of race is a principle upheld by white supremacists and exists as the backbone of racial hierarchies. Theft of identity is a systemic pattern of oppression that has impacted Indigenous tribes historically (Herndon, 2018). Thus, there is continuity in the reification of inheritance in modern genetic study.

We live in an era of commercialized genomics, where ancestral profiles can be delivered to our doorstep. Genetic ancestry complicates ideas of identity, heritage, and race in many ways far more than it clarifies. While it can yield a "discovery", a restoration of roots, a confirmation of origin, a license for citizenship depending on who, where, and when, lurking beneath ancestral science are assumptions about race, nation, and culture framed by an imperialist and racist past.

CHAPTER ONE

A Brief Introduction to Modern Genomics

Following the sequencing of the human genome, geneticists derived two striking observations. First, about 99% of the genome proved identical across all people (Batai & Kittles, 2013; Fine et al., 2005; Rosenberg et al., 2002). Second, of the remaining one percent, ~87-95% of sequential differences arise from intrapopulation genetic variation, or the variability within groups (Batai & Kittles, 2013; Fine et al., 2005; Rosenberg et al., 2002). Though several authors cite slightly different percentages, they collectively describe interpopulation genetic variation within the approximate window of 5-13% (Batai & Kittles, 2013; Fine et al., 2005; Holsinger & Weir, 2009; Rosenberg et al., 2002). Essentially, under the circumstance that ancestry is a component factor of human genetic variation, differences between ancestral lineages could only be identifiable in 5-13% of 1% of the human genome. Genetic distinctions between human populations extrapolate from a small percentage of one percent of detectable difference. The remarkable similarities of our genomes swiftly undermined notions of biological determinism, countering a heinous past of erroneously exploiting race in science; however, geneticists quickly pivoted toward investigations of population structure[1], pursuing the promise of ancestry-linked disease risk theoretically encoded within genome-wide variation.

This focus of genomics on ancestry and population, while maintaining risks of misapplication and misinterpretation, has advanced considerably, translating hypothetical approaches surmised in the mid-1900s into numerous genotyping methodologies by 2005 and

[1] Population structure describes patterns of genetic variation within and across populations (Henn et al., 2010; Rosenberg et al., 2002).

new computational metrics throughout the 2010s. Several moving parts form the technical scaffolding of modern genomics, including independent reference databases, analytical techniques, statistical models and correction factors, and genetic marker panels as well as post-study validation through replication and meta-analysis. The present chapter will overview the ideas and mechanisms behind modern genomics, deciphering how geneticists conceptualize genetic ancestry, demarcate human evolution, and apply population genetics to real people.

When it comes to the science of genetics, most of us start from, and end with, a basic understanding of Mendel's laws. Gregor Mendel, a 19th century scientist, modeled laws of inheritance through his famed pea plant experiments. He revealed the existence of discrete, heritable elements that govern how organisms pass down visible traits to offspring (Westerlund & Fairbanks, 2010). Most notably, he discovered that these elements, or alleles, segregate and randomly reassort during the genesis of reproductive cells into a set of predictable outcomes (Westerlund & Fairbanks, 2010). Although his work predates chromosomal theory, the discovery of DNA structure and meiotic function further revealed his laws of inheritance at work within the cell (Gayon, 2016). And so the story goes—combinations of alleles form genotypes that express observable traits through the transcription and translation of DNA into proteins and other molecules. Together, these cellular products form our metabolic pathways, construct our tissues, regulate our physiology, and interdependently sustain life. Punnett squares, pedigrees, and other tools accurately model Mendelian traits, enabling scientists to infer parental genotypes or project the expected genotypic and phenotypic ratios of a subsequent generation. Today, students continue to visualize Mendelian inheritance by crafting Punnett squares and tracing the strangely succinct patterns of allele segregation and random assortment to predict simulated genotypic combinations; however, simple inheritance is limited in scope, unable to fully encapsulate the

nuances of heredity at a chromosomal level. In fact, it is the deviations from Mendel's laws that actually proved crucial for the development of modern analytical techniques and theoretical underpinnings.

After Mendel's pea plants, geneticists soon noticed unexpected patterns in genotypic ratios that seemed to challenge Mendel's laws; this phenomenon would eventually be explained by the concept of genetic linkage, which developed alongside the chromosomal theory of inheritance. Chromosomal theory enabled geneticists to figure out that, on a molecular level, genes comprise fixed stretches of DNA, called genetic loci, on the chromosomes. During meiosis, homologous chromosomes trade genetic material in a process called crossing over, or recombination, yielding new combinations of alleles from parental building blocks that will be inherited by the offspring (Lobo & Shaw, 2008). Crossing over typically transfers each gene independently and randomly; however, when two genetic loci are physically proximal on the same chromosome, there is a probability that the genes will be inherited together, or genetically linked (Lobo & Shaw, 2008). Thus, genetic linkage is a function of gene proximity that can skew the randomness of crossover events as well as the expected allelic combinations of simple Mendelian inheritance. Grasping the theory of genetic linkage, geneticists developed linkage analysis, a statistical method for mapping inheritance and discerning the chromosomal location of genes (Smith & O'Brien, 2005). To do this, geneticists look to genetic markers, which are specific DNA sequences that occupy known locations throughout the human genome. This technique leverages the co-segregation of genetic markers with traits-of-interest to trace the chromosomal region that contains the associated gene(s). By analyzing co-segregation, geneticists can associate a trait-of-interest with a genetic marker and search nearby for causal

genes (Darvasi, 2005). The concept of linkage, and specifically the linkage of genetic markers with gene variants, form the conceptual basis of most modern analytical techniques in genetics.

Genetic markers are essential tools in modern genomics. Both convenient and naturally-occurring, these microscopic signals proliferate throughout the human genome. Genetic markers are highly polymorphic, meaning they can take on variable forms in different individuals (NHGRI, n.d., para. 1). Different classes of polymorphisms occur throughout the genome, and new forms typically emerge through benign mutations.

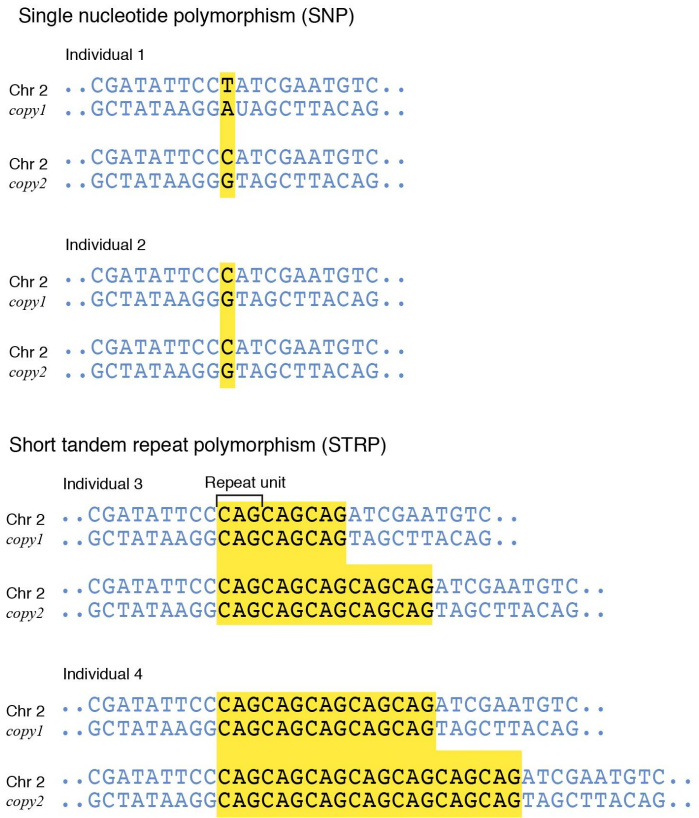


Figure 1. Diagram displaying single nucleotide polymorphism and short tandem repeat polymorphism marker classes. Individuals might differ in the nucleotide base (A, T, C, or G) present in an SNP or the number of repeats in an STRP. Reprinted from the National Human

Genome Research Institute (NHGRI), Retrieved April 23 2020 from https://www.genome.gov/genetics-glossary/Polymorphism

Microsatellites, or short tandem repeats, result from DNA replication malfunction, disproportionate recombination, or other mutations that either disrupt or elongate sequential repeats (Gilson & Tassis, 2007). Conversely, single-nucleotide polymorphisms (SNP) are point mutations that produce variants via single base-pair alterations in DNA (Bush & Moore, 2012; Gilson & Tassis, 2007). New microsatellites recur more regularly than new SNPs, but the total number of SNP variants exceeds those of microsatellites by ~1000x, making them the most common type of human genetic variation (Bush & Moore, 2012; Gilson & Tassis, 2007). In a group of individuals, the majority might possess one form of a microsatellite or SNP while a small percentage might possess a detectably different form. Combine this natural phenomenon with knowledge of their chromosomal locations and genetic markers prove to be ideal locators for undiscovered genes; however, because vast multitudes of polymorphisms occur across the genome, genotyping large numbers of them can become challenging and costly. Some geneticists might employ haplotypes, which are extended regions of DNA that are inherited as a unit; these regions contain groups of polymorphisms, such as several SNPs that display patterns of co-inheritance (Gilson & Tassis, 2007, Smith & O'Brien 2005). Haplotypes allow geneticists to follow sets of linked genetic markers without genotyping them individually (Smith & O'Brien, 2005). Ultimately, different marker classes are advantageous for different modalities of analysis, but genetic markers overall are an integral part of modern genomic analysis.

The ability to locate, mark, and correlate both known and unknown segments of the genome opened the doors to inferring genetic ancestry. Genetic ancestry generally defines as

"the origin or background of our genomes " (Pardo-Seco et al., 2014, p.1), where the genome is an amalgam of discrete segments that each possess its own ancestral origin and evolutionary history. Geneticists cite multiple purposes for the investigation of genetic ancestry, including biomedical research, forensics analysis, and the general study of human evolutionary history. Each focus rests on the premise that demographic events over the course of history produced detectable allele frequency differences in specific cohorts of people and have been passed down over generations, comprising the estimated margin of human interpopulation differentiation. In essence, various evolutionary processes may have been unintended byproducts of our demographic histories, although in much more complex and interweaving ways than other species. For example, genetic drift describes events that randomly restructure the allele frequencies throughout the entire genome (Elhaik, 2012; Holsinger & Weir, 2009). These allele frequency changes can occur through various means, such as a founder effect, whereby a subpopulation migrates out of its original population and reestablishes a geographically or otherwise distinct community, or a bottleneck event, during which natural disasters, famines, plagues, or other sweeping external impetuses result in the survival of a genetically random cohort; if the new cohort is small enough, certain allelic forms may become fixed or completely lost simply due to chance (Kliman et al., 2008). Whenever a group of individuals become sequestered from their original communities, whether voluntarily or forcibly, their gene pool will contain new proportions of randomly up- and down-regulated allele frequencies.

Geneticists have been trying to interpret and visualize the small, estimated range of human interpopulation differentiation for the past several decades. They estimate 7.6% of all allelic differences detected among geographic regions to be circumscribed to a single region, most of which are rare variants typifying relative frequencies of 1.0% within each region

(Rosenberg et al., 2002). Amidst the early findings of population structure research, Rosenberg et al. (2002) make the crucial distinction that "most studies of human variation begin by sampling from predefined 'populations.' These populations are usually defined on the basis of culture or geography and might not reflect underlying genetic relationships" (p.2381). Chronology is pivotal in genetic studies that correlate populations with genetic variation, and several of the limitations discussed in subsequent chapters will center chronological paradoxes. Many studies delineate populations according to descriptive metrics, such as geographic or demographic characteristics, before they collect DNA samples from the individuals within the populations. Studies of population structure convey a conspicuous endeavor to systematize humans according to empirical genetic boundaries, so the context of their methodological structure is highly important. Geneticists have also made decisions regarding which processes have most likely impacted human population structure, which diverge from models pertaining to other organisms.

In order for interpopulation genetic ancestry to be relevant, historic human groups must have met certain evolutionary parameters. Namely, human "population" differentiation occurs when a community subsists within reproductive barriers for a significant span of generational time (Henn et al., 2010). Population structure research typically establishes relevant boundaries according to broad geographical regions. Some geneticists have honed in on fine-scale population structure, which studies multi-generational pedigrees stemming from endogamous familial structures (Henn et al., 2010). Broadly, "endogamy" refers to the convention of marrying only within one's community affiliation. In essence, geneticists might argue that, while geographic boundaries are likely influential in human population structure, sociocultural factors that surround concepts of family, marriage, and community may also play a role. For example,

the Indian Genome Variation database (IGVdb) project (2005) describes the endogamous marriage practices within the castes of the majority Hindu population in India. Some researchers have cited evidence for population structure using specific case studies. Henn et al. (2010) reference the Tibetan peoples in the Qinghai Tibet Plateau region, as the gene pool of this community shows higher frequencies of rare alleles (which actually appear to be adapted to high-altitude hypoxic conditions) compared to the surrounding lowland Han Chinese individuals (Henn, 2010). Henn et al. (2010) also relay that "many models predict that only a limited amount of migration is required to largely eliminate differences in population frequencies" (p. R224), complicating the role of geographic or endogamous boundaries. While conceptual theories and case studies might reflect how diverging ancestral histories may surface in genetic analyses, there are significant challenges to extrapolating these patterns across all humankind.

Quantifying population structure hinges on the range and resolution of computational methods. Early studies attempted different kinds of associations, such as between gene trees and language groupings, as linguistic and geographical distinctions appeared to correlate (Henn et al., 2010). A later development was the use of mitochondrial DNA (mtDNA) and Y-chromosome DNA rather than autosomal DNA (Henn et al., 2010). Some ancestry tests still leverage mtDNA and Y-chromosome DNA, both of which are passed down largely unchanged from maternal and paternal lines respectively. Geneticists across focuses have extensively employed polymorphic genetic markers, including autosomal microsatellites, SNPs, and haplotypes, trying to construct ancestry informative marker (AIM)[2] panels that are more informative when used in specific

[2] AIMs are population-specific genetic markers that comprise polymorphisms of known locations and relative population frequencies (Batai & Kittles, 2013). Geneticists use statistics to compute and publish the relative "informativeness" of AIM panels in different populations. Pardo-Seco et al. (2014) argue via their AIM evaluation across three continental groups that the number of AIMs in a panel is more important than their informativeness for estimating genetic ancestry. Ideally, the designation of AIMs should facilitate the ability of geneticists to account for potential population structure within association studies.

populations (Batai & Kittles, 2013; Henn et al., 2010). The development of genetic marker panels has coincided with advancing statistical methods used to compare their relative performance. To add another layer of complexity, the selection of statistical metrics proves a crucial component of methodology and guides the interpretation of data[3]. Essentially, robust analytical methods, genetic marker panels, reduced costs, and other advancements in the field have allowed geneticists to investigate the minute window of interpopulation difference with more precision than ever before, fueling the study and application of genetic ancestry.

Biomedical research is one of the contexts where genetic ancestry is most prominently applied. Geneticists believe that studying interpopulation difference could shed light on differential risk and enhance preventative medicine. Linkage analysis provides insights into disease risk within family pedigrees, proving useful for simple, Mendelian patterns of inheritance and monogenic pathologies[4] (Smith & O'Brien, 2005). Yet, this analytical technique cannot capture the complexity of common diseases and their underlying mechanisms of inheritance (Bush & Moore, 2012; Smith & O'Brien, 2005). Instead, geneticists implement analytical tools tailored to population genetics, including genome-wide association studies (GWAS) and

[3] Geneticists have developed statistics and estimators to describe genetic marker information content as well as analyze genome maps. F-statistics, which are among the most pertinent metrics of interpopulation variation, measure marker information and quantify interpopulation differentiation. Developed in the early 1900s, Wright's F-statistics comprise an umbrella of fixation indices that capture heterozygosity, or the level of genetic variability, within different population stratifications, such as at the level of the individual, subpopulation, or total population. (Elhaik, 2012; Holsinger & Weir, 2009). While the three constituent F-statistics (F_{ST}, F_{IS}, and F_{IT}) all describe populations on the basis of variance, or deviation about a mean value, each designates a particular ratio (Holsinger & Weir, 2009). For example, the subscript of F_{ST} denotes "subpopulations within total" while that of F_{IS} describes "individuals within populations" (Holsinger & Weir, 2009, p.5). F_{ST} is markedly used in population genetics to quantify interpopulation allele frequency variance, indirectly reflecting the degree of genetic similarity among individuals within a population (Elhaik, 2012; Holsinger & Weir, 2009). When interpreting Fst statistics, a low Fst value means that two populations have intersecting gene pools while a high Fst indicates that two populations are relatively isolated from each other. AIMs typically display notably high F_{ST} values, but this value fluctuates depending on genetic marker type (e.g. microsatellites versus SNPs) (Betai & Kittles, 2013). In their patent publication on best practices for AIM selection, Frudakis and Shriver (2004) designated marker panels that display significant frequency differences ($F_{ST} > 0.4$) between populations as informative of ancestry.

[4] Monogenic refers to traits expressed by a single gene rather than multiple genetic loci (Smith, 2005).

mapping by admixture linkage disequilibrium (MALD or "admixture mapping"). These methods have been developed with a keen focus on inferring disease-causing gene variants by correlating chromosomal regions and disease prevalence. Both mapping techniques have also investigated population structure, pursuing the notion that ancestry could be an important contributor to the inheritance of certain disease variants, but MALD designates more specific focus to differential ancestry.

Initially a theoretical idea, MALD, or admixture mapping, has become widely used to genetically compare ancestries. Geneticists believe certain groups of people, or "admixed populations", to represent the recent convergence of two or more ancestries, termed "parent" or "progenitor" populations. Assuming that parent populations have genetically differentiated from each other overtime, geneticists will assign genomic segments in admixed individuals to specific ancestries using the concept of linkage disequilibrium (LD)[5] (Darvasi, 2005). In an applied setting, geneticists believe that shorter LD blocks exist among Pan-African individuals, theorizing that humankind originated in modern-day continental Africa (Bush & Moore, 2012; Zheng-Bradley & Flicek, 2017). They describe the dispersal of groups to Europe, Asia, and America as created by founder effects that altered population size and generational age, resulting in longer LD regions in descendent individuals today (Bush & Moore, 2012; Zheng-Bradley & Flicek, 2017). In simple terms, MALD uses the concepts surrounding LD to help identify which chromosomal regions come from which ancestries, expecting to find different sizes of LD blocks

[5] Linkage disequilibrium (LD) is a function of the linkage between alleles across the genome; when a population is in high LD, alleles are frequently linked when passed from generation to generation (Bush & Moore, 2012). During the process of LD, multiple instances of recombination occur with each new generation, fragmenting chromosomal regions (LD blocks) over time (Bush & Moore, 2012). The splitting of these chromosomal regions amplifies as the number of generations increases (Bush & Moore, 2012). If a population remains relatively fixed in size and sustains random mating patterns, crossover events will continue to split apart contiguous chromosomal segments containing linked alleles until every allele within the gene pool of the population reaches linkage equilibrium, or essentially become unlinked (Bush & Moore, 2012).

depending upon the degree of admixture in a sample population. Geneticists use MALD to identify polymorphism proportions across human populations, help construct ancestry informative genetic marker panels (AIMs), and characterize genetic differences between ancestries.

In a biomedical context, admixture mapping is designed to extract evidence for "differential risk by ancestry" (Shriner, 2013, p. 1.23.2). Using admixture mapping, geneticists will try to discern whether an allele frequency difference between two ancestral groups associates with disease prevalence. Detecting "differential risk by ancestry" requires the following criteria to be met: there is a measurable difference in the prevalence of a disease-causing allele between parent populations, admixture has occurred over the course of at least two generations, and there is a usable marker set that can differentiate chromosomal regions from each parent population (Smith & O'Brien, 2005).

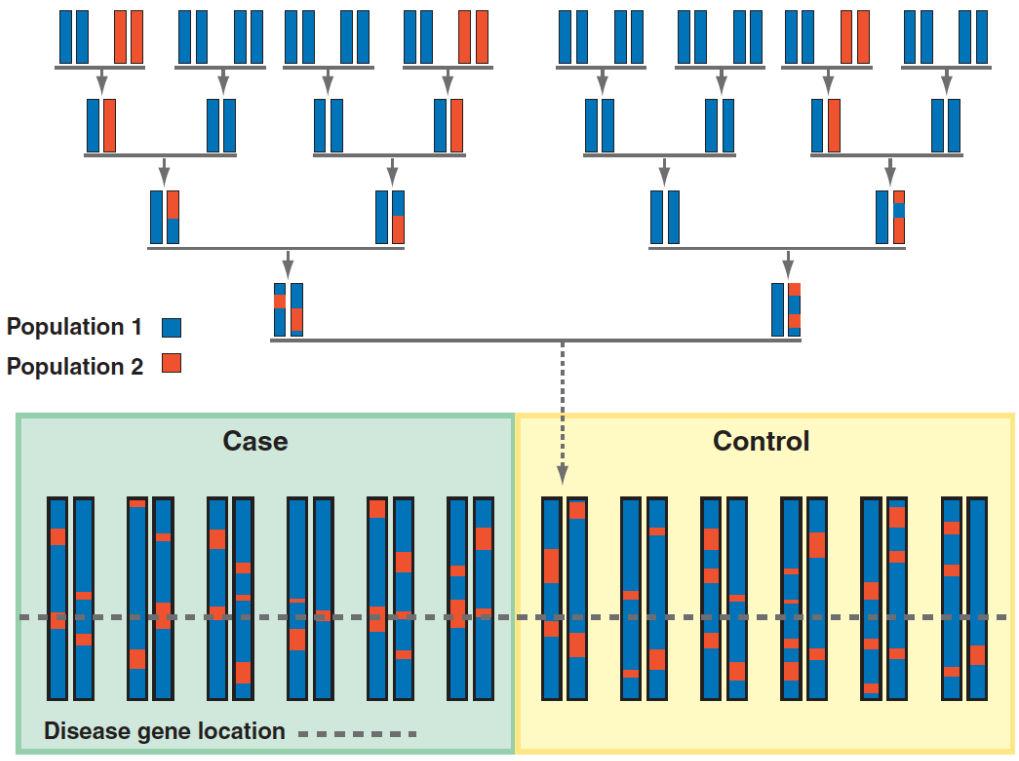


Figure 2. Diagram comparing the same chromosome across several individuals in an admixed population, separated into case (carrying the disease) and control groups. The dotted line represents the disease-causing locus. For admixture mapping to work, the parent populations (red and blue) must have displayed different allele frequencies at the disease-causing locus. This diagram suggests that Population 2 (red) has a higher prevalence of the disease allele, leading geneticists to conclude that Population 2 ancestry possesses a higher risk for the disease-of-interest and carries a higher frequency of the causal allele. Reprinted from "The Beauty of Admixture", by Darvasi & Shifman, 2005, *Nature Genetics*, *37*(2), 118.

In an ideal scenario, geneticists will collect DNA samples from a group of disease-carrying individuals and a group of healthy individuals within an admixed population (Smith & O'Brien, 2005). Then, they will use a highly-informative genetic marker panel[6] to examine chromosomal segments across every individual (Smith & O'Brien, 2005). Finally, by comparing across chromosomal segments, researchers will try to detect a genetic locus in the disease case group that displays a disproportionate prevalence of ancestral DNA from one or the other parent population (Smith & O'Brien, 2005). The result is an association between a chromosomal region or genetic locus and an ancestral group, which cannot alone indicate whether the locus actually causes the disease. Geneticists must use (or have previously used) GWAS to discern whether a given genetic locus is linked to disease risk. In joint testing methods, GWAS can be used to increase the resolution of admixture signals (Shriner, 2013). For example, several studies have attempted to locate contributory genes for heart disease and related risk factors, such as blood

[6] For MALD, it is recommended that a marker panel has been previously tested and displays frequency differences above 60% between parent populations (Smith & O'Brien, 2005). Markers that are considered fully informative, which are 100% prevalent in one parent population and 0% in the other, exist in an ideal situation; yet, the human genome contains a minute number of fully informative markers (Shriner, 2013).

pressure and cholesterol levels, as well as evaluate and reevaluate the role of ancestry by controlling for non-genetic factors (Batai & Kittles, 2013; Bush & Moore 2012; Zhu et al., 2011). So, admixture mapping correlates phenotype[7] to ancestry while association studies correlate phenotype with causal genes.

Similarly to the other methods, geneticists using GWAS start by discerning associations between chromosomal regions and disease prevalence. Different disease types inform the methodology and overall investigation of genetic risk factors. A disease can either be considered a binary variable, described in terms of case versus control, or a quantitative variable that arises from continuous, quantifiable factors (Shriner, 2013). In circumstances where a single genetic variant correlates directly with the disease incidence, the variant is often classified as a case-control variable. Mendelian diseases, which are typically rare, are simplistic in that researchers can reasonably predict genotype based on a given phenotypic value (Shriner, 2103). For example, Cystic Fibrosis is a Mendelian, autosomal recessive disorder, which means that the combination of two recessive alleles at a single genetic locus on an autosomal chromosome is an all-or-none determinant of whether an individual inherits the disease (Chial, 2008). Complex diseases are generally more commonly prevalent but also may include certain rare disorders. The leading hypothesis among disease susceptibility research purports that common genetic variants, rather than uncommon alleles, are likely contributory to common diseases (Bush & Moore, 2012). These disease phenotypes possess more convoluted inheritance patterns, as they might involve a few or many genes as well as classify as multifactorial[8].

[7] Broadly, phenotype refers to an expressed trait that is observable in an individual (Shriner, 2013). Phenotypes can include visible attributes, physiological characteristics, or any trait that arises from the interaction genes and the environment.

[8] Multifactorial diseases arise from the combined effect of multiple genes, environmental factors, and gene-environment interactions (Shriner, 2013). Multifactorial traits may also be continuous, meaning they display phenotypic gradations rather than binary incidence (Lobo & Shaw, 2008).

Complementary case-control and quantitative study structures exist, both rendering important distinctions and limitations. When studying a case-control model, geneticists do not test for associations between a potentially causal locus and partitioned risk factors beyond the overarching disease-incidence (Bush & Moore, 2012). Associations are solely made between case and causal variant(s). In quantitative studies, geneticists have more specifically defined phenotypic contributors to disease (Bush & Moore, 2013). For example, under the heart disease umbrella exist several observable phenotypic risk factors, such as the role of HDL and LDL levels (Bush & Moore, 2012). Thus, quantitative studies might specifically search for causal variants that associate with HDL and LDL levels and other risk factors rather than only with the case of heart disease (Bush & Moore, 2012). Overall, case-control studies, while informative for simple patterns of inheritance, may overlook the branching complexity or more common diseases. As a result, quantitative and case-control studies have vastly different statistical power and measurement errors (Bush & Moore, 2012). While genome-wide association studies into genetic disease susceptibility offer promise into furthering public health research, they offer associative values rather than discretely conclusive data.

The clinical promise of genomics incentivizes modern geneticists but poses significant challenges.Varying correlative and statistical measures offer a lens into the role of inheritance and genetic-linkage in human diseases; however, genetic analysis cannot provide conclusive results, as it primarily functions as a probabilistic science. It is crucially important to understand the relationship between disease-incidence and the interactivity of underlying genetic loci. In effect, we must understand the necessary assumptions and distinctions within the phenotype-genotype relationship. An observable trait may possess a genetic underpinning, but several

interrelated factors intercept a direct one-to-one relationship between genotypic expression and observable phenotypes (Shriner, 2013). Although a non-scientific audience may conceptualize genetics according to Mendel's simple patterns and rare disease variants, modern genomics has uncovered important complexities. Most common human diseases, such as coronary artery disease, are multifactorial. While Mendel proposed that genetic elements are discrete and independent, multifactorial studies have indicated that genes can regulate the expression of other genes through their allelic pairs or the molecular products they encode (Lobo & Shaw, 2008). Certain disease phenotypes may observably correlate with different populations, but phenotypic distributions do not inherently reveal an equivalent genotypic distribution within the underlying genetic structure (Shriner, 2013). In parallel, analytical methods that attempt to associate traits or disease types with a causal genetic locus cannot assume that risk factors distribute evenly across a given genomic region (Shriner, 2013). Phenotype is an *emergent phenomenon*, a whole greater than the sum of its components. Regardless of associative value, geneticists cannot assume the relative contribution of genetic versus environmental factors on differential risk by ancestry (Shriner, 2013), a limitation that will be closely examined in Chapter 5. These assumptions help inform the degree of information that genome maps provide. Geneticists must either control for multiple variables or clearly cite limitations when conjecturing the root of phenotypic interpopulation differences.

Genetics is a tricky probabilistic science that relies upon sound methodology and statistical power. It speaks a language of comparisons and associations rather than strict causality, which can be easily misinterpreted. When referring to genetic ancestry, geneticists have a central role in refracting how we organize and ultimately "see" the data, as gene pools exist within a constant and dynamic continuum of cross-generational change. Several analytical

limitations can also impede study design. In GWAS or MALD, genetic markers can show a false-positive association with a disease case (Bush & Moore, 2012). Admixture mapping requires population-specific marker panels, and researchers can over- or under-estimate the extent of ancestry in a chromosomal region if they miscalculate allele frequencies within one of the parental populations (Smith & O'Brien, 2005) Furthermore, MALD rests on a slew of assumptions that greatly diminishes mapping power if they are not met. If inclusion criteria are unsound, sampling effects can predispose error when the genotypic makeup of DNA samples have their own unique allelic proportions that do not reflect the population from whom they were collected (Holsinger & Weir, 2009). Nonetheless, the field has seen strides in technical capacity but faltered in the ability to consistently and intentionally define what a human population is. Further, the complexity of study design and probabilistic nuance of findings are often glazed over amidst the translation of study to its applicable context. Geneticists are not devoting attention to the implications of their rhetoric surrounding human difference, and the dense pool of primary literature fails to provide needed transparency. Modern genomics, and science at large, has emerged from an entrenched history of malintention and maluse, whereby concepts of heredity were leveraged to argue natural schemas of superiority and inferiority based on perceived and imagined human differences. Chapter 2 will delve into the scaffolding of scientific racism while Chapter 3 will trace lines of continuity through genetic discourses today.

CHAPTER TWO

Imagining and Reimagining

A History of Heredity, Science and The Race Concept

While today's field of genomics fundamentally differs from pseudo-scientific ideas of purity, creationism, and species fixity, scientific racism still persists, using new arguments and analogies to support deterministic fallacies. While the ever-increasing and intriguing intricacy of genetic study has made simplistic fabrications of innate racial difference progressively more difficult to argue, people continue to use genetics to support binary understandings of inheritance and make reductive claims of morality, intelligence, or aptitude as intergroup biological entities. It is vital to be aware of this context and the potential for advancements in the field or its theories to be falsely contorted. For instance, Michael Levin's book, *Why Race Matters*, makes the philosophical argument that "if breeds of dog may differ in intelligence and temperament, there seems to be no reason evolution could not have differentiated human groups along similar lines" (as cited in Garrod, 2006, p.56). Another book, entitled *Race: The Reality of Human Differences*, tries to procure social and scientific evidence of discrete races (as cited in Garrod, 2006). Authors Vincent Sarich and Frank Miele make the same dog breed analogy to postulate intra-species variation (as cited in Garrod, 2006). Alongside the glaringly obvious realities that dogs were intentionally bred by humans and are not cognitively or behaviorally analogous to humankind, these authors make the fundamental fallacy of forgetting that traits are inherited non-contiguously rather than "as a group" (Garrod, 2006, p.56). Scientists and geneticists have an intellectual and moral responsibility to prevent modern genetic research, and human evolutionary sub-disciplines, such as admixture mapping, to be easily manipulated as corroborating evidence

for racist ideology. If researchers are not actively contending, or informed by, racialization processes, they risk predisposing modern reinventions of biological race and scientific racism; additionally, study designs implicitly devoid of the perspective that racism manifests systemically as a political structure and "socio-economic order" (Azarmandi, 2017, p.22) will not be sensitive to the real consequences faced by the implicated communities. The idea that intelligence and temperament evolves "along similar lines" is fundamentally wrong but demonstrates an important instance of continuity. The presumptuous nature of these ideas is neither unprecedented nor happenstance but is a perspective that reverberates against an all too familiar past. Why is continuity important for us to consider? To shed light on such a complex cornerstone of collective consciousness, we must examine the history of race in the scientific imagination.

2.1 The Social Process of Racialization

Several researchers have attempted to dissect the process of racialization and the roots of racism in the Western gaze by studying concepts of group difference throughout time, from ancient societies through recent centuries; however, race as it exists today is fundamentally tied to colonialism and slavery, which catalyzed the social and political formation of the race concept by mobilizing systems of oppression and structural injustices (Azarmandi, 2017). Racialization refers to the sociological understanding of race as a social process of organization rather than an inherent property or preexisting entity. This idea makes sense considering that race cannot be coherently schematized, as it changes across spans of time and space as well as in relation to self-concept and extrinsic perception. Race has historically been delineated according to perceived physical attributes, which were then attributed to immutable group traits and associated with cognitive or moral dispositions (Crichlow, 1993). Given that race is constructed,

affixed to status, relative to time and location, and causal of social realities, sociologists have come to understand individuals as racialized rather than race existing as a necessary truth or biological given (Crichlow, 1993). Race is thus a socio-historic rather than biological concept.

Appropriations of power and delimitations of status generate superior-inferior racial stratifications within a society (Crichlow, 1993). An important element of racialization is the creation of a default "normal" that designates one group as implicitly and explicitly representative of humanity, or some peak of humanity. In America, whiteness was positioned as the societal norm and formed as a racial identity by comparison to its deviations, perceived by white Americans to be displaced Native Americans, enslaved African peoples, and different groups of immigrants over time (Guess, 2006). What, or who, a society deems abnormal becomes "othered", facing increased exposure to stigmatization, or a "disapproval for nonconformity" (Crichlow, 1993, para. 38). Whiteness operates as the norm but is also racialized, as social, political, and legal processes enable individuals to be classified as, access, and maintain being white. Since racialization has embedded structural consequences and social realities, omissions of racial categories do not merely expose race as a social construct but engage in a collective amnesia of the socio-historic processes of racialization, rendering whiteness a "racialized form of privilege" (Azarmandi, 2017, p.28). We will see discourses fixate on explaining divergence from the norm, both a response to and ingredient of the racialization process itself.

On the other hand, ethnicity, while also socially constructed, characterizes how one identifies with a collective based on shared values, cultures, institutions, as well as language or religion (Crichlow, 1993; Deng, 1997). Ancestry, and a sense of heritage, often also intertwines with belonging. An important facet of ethnicity is shared consciousness (Crichlow, 1993), or

collective memory. Forming and sustaining collective memory across familial or geographic boundaries fosters intrinsic feelings of kinship with other people, even those we do not personally know. The socio-historic processes of racialization and ethnic formation have been at interplay within our world systems. Certain communities that are considered ethnicities or nationalities today have at one point, in different places, been considered races (Crichlow, 1993). National signifiers (i.e. Irish) and religious-cultural markers (i.e. Jewish) have historically been used as racial designations (Crichlow, 1993; Azarmandi, 2017). Past forms of "othering" and their continuity today are not equivalent across ethnic groups, as the historical oppression of Jews is not level with the racial conceptualization of all other "ethnic whites" (Azarmandi, 2017, p26), and only certain groups have been allowed white status over time; these differences relate to the continued role of coloniality in modern racialization. Several concurrent perspectives on racialization pervade both historical and modern contexts in a continuous evolution of process, incentive, and materialization. Thinking about radicalization abstractly will help us critically analyze the race concept and its historical continuity in the following overview of scientific racism.

2.2 An Abridged History of Scientific Racism

Ensuing debate surrounds the intellectual foundation of racism in the colonial, neocolonial, and modern ages, leading some researchers to survey concepts of group difference through the years of the common era and beyond; however, pre-colonial sources did not conceive race but were mobilized later on in both the formation of European identity and justification of colonial exploitation. Greek and Roman ideas became cornerstones of philosophy and society in Europe. These ideas carried forward through the medieval period and the Enlightenment, eventually pervading into the colonial era. The ancient Greeks' outgroup

designations centered on culture and politics rather than physical appearance, viewing non-Greek culture as barbaric; but, Greek society allowed persons to "shed" inferior status by assimilating with the *polis* (Graves, 2003; Yudell, 2011). We see an opaquely similar notion appear in the colonial concept of modernity, viewing colonized land as uncivilized and tamable only by Europeans, but racial status was seen as innate and could bar access to assimilation[9]. Imaginations of inherent European modernity were also embedded in environmental determinism, or the belief that the environment shapes the disposition and ability of entire groups, which we will see in 19th century naturalist characterizations of Latin America. But, the roots of environmental determinism are detectable in Hippocrates's argument that soil infertility in Greek civilizations produced a self-reliant and superior people while tropical fertility and abundance "led to softness and lack of war spirit", specifically of the Asiatics (Graves, 2003, p.17). Conversely, the Romans cast northern Europeans as intellectually inferior due to cold and humidity (Graves, 2013). We can also look to the Romans for pre-scientific suppositions of interpopulation variation. As early as the fourth century AD, Roman Emperor Julian the Apostate observed that the physical appearance of human bodies varied across geographical locales and conjectured that temperament and cognitive capacity likely did as well (Graves, 2013). He proceeded to classify several civilizations according to psychological attributes, inherited by the different gods who fathered them (Graves, 2013). We will see this notion of group genealogy and contiguous inheritance resurface as biological concepts in the 19th century and beyond. Finally, Graves (2013) argues that certain key Greco-Roman concepts, such as Aristotle's belief that some people were born natural slaves versus natural rulers, were reapplied later on to justify racist hierarchies. Aristotle's natural slave concept would eventually be cited

[9] Nonetheless, assimilation is also a form of erasure and social control that was forcibly leveraged against Indigenous peoples, such as in 19th century America (Tallbear, 2013).

by Spanish humanist Juan Gínes de Sepúlveda during the Valladolid debates in 1500-1 to rationalize slavery within the colonial project (Brunstetter & Zartner, 2011). Sources of perceived philosophical credibility were retroactively tapped as the race concept materialized during colonial exploitation and intersected with scientific development

Concepts of group difference as tied to blood and lineage amplified during the Middle Ages, specifically within the consolidation of anti-Semitism. In Medieval Europe, virulent disdain of the Jewish faith adopted a new framework of "blood kinship" and evolved from the hatred of a religion into "hatred of a people" (Yudell, 2011, p.15). Gross constructions of Jewish people as physically, cognitively, and morally derelict legitimized widespread persecution (Graves, 2013). Such constructs were often sensationalized caricatures, some rooted in superstitious hysteria, describing Jewish people as having horns, emitting foul smells, or ritualistically murdering youths or Christians for their uncompromised blood (Graves, 2013). Calls to action by various Catholic popes during the years 1000-1300 prompted the genocide of Jewish communities in Worms[10] by Crusaders as well as requirements that Jews wear identifying badges and live in ghettos (Graves, 2013). There is obvious contintuity between these oppressive policies and those reinvented during the Nazi regime; understanding that history is prone to repeat, we can analyze past constructions of group difference and positionality with attentiveness to trends of the present.

We can look to colonialism to observe how Eurocentric imaginations of race acquire a notably naturalistic tone. The "Age of Discovery", spanning from the 15th to 17th centuries, saw more frequent in-person encounters between European, Indigenous and African peoples and cultures (Graves, 2013). "Discovery" of the Americas piqued European colonizers' interest in

[10] Located in modern-day Germany

examining “uncharted” territory, enticed by the prospect of novel research settings and the prestige of exploration. With European fixation on an imagined “New World" came conquest and colonization. Crimes against humanity advanced under the guise of enterprise and became business as usual, embedded through a promulgation of ideas that cultivated popular notions of exoticism and race in the Americas. Naturalist inquiry and colonial travel writing proved to be arms of the colonial project, playing a significant role in justifying the subjugation of human beings by reinforcing a superior European image through contrast with a designated “other”. The slave trade pervaded conquest of the Americas, resulting in devastating losses of lives as well as attacks on culture, severing ties between families and communities. Graves (2013) denotes how translocating slavery to American colonies built an evironment of self-sustaining racism:

> The slave trade not only brought together populations that previously had been geographically separated but also brought them together under conditions of manifest social inequality. That is, phenotypic characteristics were used to symbolize social status…. The absence of well-validated theories of heredity meant that no one really understood which features of human beings were innate…. (p.30)

Oppressive systems manipulate different theories, whether religious, philosophical, or eventually scientific, to naturalize oppression, assuming a naturalistic lens as phenotypes were racialized and assigned social meaning by colonizers in the Americas. Although early colonizers did not understand concepts of heredity, they recurrently distorted successive scientific theories of inheritance and evolution to argue inherent inferiority. Their ideas circulated widely and fueled the racialization process throughout European colonies and homelands. We can look to

colonization and exploration[11] in Latin America as a case study for the dissemination of ideas on European superiority and normativity.

European colonizers positioned political and economic interests under the guise of conquest and progress, forming their “empires" atop the infrastructure of the already existing, culturally rich, and sophisticated civilizations of Latin America. In order to both justify the marginalization of non-white communities as well as ensure political and economic strategems, colonizers methodically constructed evolving versions of racial hierarchies through naturalist discouses as travel to Latin America was popularized in the 18th and 19th centuries. Stepan (2001) proposes the key theory that travel accounts proved integral to the formation of European identity via the “othering” of the Tropics of Latin America. This theory reflects the idea that racialization simultaneously normalizes and “others” and nods to travel writing as the primary modality that shaped the race concept during this time period (Stepan, 2001). Naturalists were keenly descriptive, conveying reports of what and who they witnessed. They were often key figures of the European Enlightenment, a pivotal intellectual campaign that promoted precepts of reason and empiricism as well as shaped Western concepts of the natural sciences; the Enlightenment thinkers set the tone for scientific naturalism, as philosophers like René Descartes (1596–1650) reimagined science as a “deductive procedure” (Vartanian, 1953, p.24) and Denis Diderot (1713–1784) held that discernable laws of matter governed all things, including “intricate organic details” (p.291). Through the medium of travelogues, white colonists rationalized entitlement to Indigenous land and African enslavement by imagining themselves to

[11] By exploration, I am referring to a 15th through 18th c. European idyllic notion, whereby Europeans investigated lands previously unknown to them and documented natural sites for the accumulation of knowledge; the ideal of exploration goes hand-in-hand with the colonial project, incentivized by the prospect of capitalizing new regions of the world through exploitation and displacement.

be racially superior; they normalized their own narratives of identity, and naturalist racial hierarchies ingrained into the collective consciousness of Europeans abroad.

The academic missions of European naturalists generated descriptive profiles of difference; these descriptions were often strikingly morphological, dehumanizing individuals as objects of scientific scrutiny. Alexander von Humboldt was among the most prominent characters of his time, a Prussian naturalist who wrote numerous romantic works on Latin American landscapes and peoples at the turn of the 19th century. He traveled throughout modern-day Venezuela, Colombia, Peru, and Ecuador, crafting a first-hand witness account of everything he observed (Stepan, 2001). Within his travelogues, Humboldt encountered various Indigenous nations, positioning his reflections on them as empirical study. For example, he delineates the Chaima Indigenous nation using markedly physical indicators, describing them as "short and thickset, with extremely broad shoulders and flat chests…" (von Humboldt, 1995, p.120). He also purports that each individual visually resembled one another, maintaining a close relatedness stemming from what he believed to be a "blood link" that endowed each a "lack of intellectual culture" (von Humboldt, 1995, p.120). By marking physical traits and perceived relatedness as indicative of the cognitive capacity of an entire group of people, Humbolt implies that there is a biological or predeterministic underpinning to the human race. Humboldt reports that the individuals he observed retained a "moral inflexibility, a stubbornness", which he believed "[characterized] the whole race from the equator to Hudson's Bay and the Strait of Magellan" (von Humboldt, 1995, p.119). Humbolt reveals his own hubris in believing he is capable of not only discerning the character, moral beliefs, and intellect of Indigenous communities but also generalizing his claims to entire populations based on his own interpretations of what ties people together. The morphological nature of naturalist discourse

helped convey intergroup difference as a perceptibly physical, and by the early 19th century, the racialization process oriented race as a physical category that could be functionally delineated according to biological differences (Murji & Solomos, 2005). Humboldt's prolific narratives, which are credited with inspiring 19th century German natural science (Reill, 2005) and prominent European scientists (i.e. Darwin) (Costa, 2009), exemplify how explicit physical description in and of itself plays a role in depersonalizing and othering groups of people as well as distancing one's own sense of self and normalcy, and it is this keen descriptiveness that characterized Enlightenment empiricism. In addition to their widespread circulation, the philosophical ideas purported by travel writers maintained a certain intellectual authority among domestic Europeans as supposed first-person accounts.

Naturalistic explanations for human difference pervaded successive scientific pools of literature. In parallel, environmental and biological determinism were imagined and reinforced throughout the 18th and 19th centuries. Byrd and Hughey (2015) summarize the emergence of biologically deterministic ideologies in the following:

> The seeds for biological determinism and racial essentialism took root in intellectual and societal discussions of race and inequality well over 500 years ago during the emerging era of colonialism. Over the centuries, the belief in the race concept—a concept that could apply to groups around the world and relate to immutable and heritage traits—was further entrenched with the development of science and medicine, particularly in the nineteenth century. (p.16)

Deterministic ideas were accepted as logical explanations for the differences Europeans observed in both the landscapes and communities they invaded. Graves (2015) defines biological determinism as the belief that social positionality is biologically ordained, emerging from the

traits inherited by advantaged versus disadvantaged peoples. Gould (as cited in Graves, 2015) proposes an antecedent iteration that biological determinism perceives socioeconomic rifts across racial groups (as well as class and gender) as emerging from innate biological dispositions, starkly contrasting the logical chronology of racialization by conjecturing that "society, in this sense, is an accurate reflection of biology" (p. 25). Biological determinism simplifies personhood and life chances as rooted in organic components, such as the size of one's brain, and eventually also heritable qualities with the advent of genetics. Byrd and Hughey (2015) define racial essentialism as the "belief that certain biological traits and social behaviors were linked and constituted the 'essence' of a certain racial group" (p.10), forming a liaison whereby biologically deterministic notions fundamentally underlie an entire social identity.

Environmental determinism relates to biological determinism and racial essentialist thinking, reasoning that differential risk by environment physically and psychologically shapes inhabitants due to the natural barriers, or lack thereof, they must overcome. For example, in 1885, a Boston journalist named Maturin M. Ballou switched from believing in the capacity, masculinity, and resources of Cubans to casting them as dearth of "self-reliance and true manhood" (as cited in Skwiot, 2010, p.53). Importantly, Ballou attributed his change of opinion to the humid climate of the Tropics, yielding over time a lavish abundance of natural foods and persistent temperature that served to dispirit its inhabitants (Skwiot, 2010). Stepan (2001) also describes popular ideas surrounding the Tropics as appearing overabundant, warm, and giving of all the natural resources humans need to survive, thus instilling idleness over industriousness. In contrast, Europeans believed the cold, temperate climate of their country fostered industriousness and collaboration due to the need to both command and shelter from their environment for

survival (Stepan, 2001). Deterministic and essentialist theories craft a direct bridge between natural science and social life, creating a philosophical scapegoat for difference and disparity.

These ideas pervaded naturalist discourse, both well-read within the public sphere and carried by European colonizers. In her work, "European Travelers and the Writing of the Brazilian Nation", Costa (2010) characterizes the archetypal 19th century traveler as a professional, one who was learned on the current pool of literature and traveled "in the service of the Academy of Sciences or the Geographical society" on accredited "missions" (p.210). Costa (2010) also notes that readership was not restricted to members or officials of missions and societies. Amidst the advancement of journalism and increasing "autonomy" of universities in the 1830s, travel writing reached progressively broader readerships among the educated, and scientific journalism developed in situ European source countries; travel writing opened the door to specialized occupations back home, including scientific journalism, university professorship, diplomacy, and government appointment (Costa, 2010). Given the intellectual foundation and ubiquity of deterministic theories, we will see the process of racialization intertwine with the fabric of scientific development as well as fuel pseudo-scientific inquiries. While genuine discoveries and theories emerged during this period, simultaneous conjectures of the human race concept took fallacious leaps from these foundations. These leaps characterize the development of scientific racism, the antithesis of the scientific process purposed with deriving foundationally unsound evidence to establish racial "types"[12] and hierarchies (Garrod, 2006). Scientifically racist theories argued that "nature, and not social forces, created divisions in society" (Garrod, 2006, p.55) and that the socioeconomic conditions, cognitive outcomes, and resource accessiblity of non-white individuals were "'scientifically ordained by Nature'" (p.55). Proponents tried to

[12] Use of the word "type" emphasizes the schematizing or taxonomic nature of naturalist and pseudo-scientific racial categories, which attempted to assign fundamental or inborn attributes to racial identities.

ascertain “evidence” to prove ideologies of biological and environmental determinism as well as further new concepts of innate superiority versus inferiority.

Understanding the broad readership as well as the role of travel writing in crafting an identity of European superiority, we can pivot to the pseudo-scientific work of 19th century traveling scientists. Race concepts of the 19th century were built upon an existing conceptual basis of kinship, ancestry, and inheritance, as even pre-Darwin philosophers focused on biblical lineages and differentiation between human groups. Polygenism emerged around 1520, when Paracelsus, a Swiss scientist and alchemist, traced European roots to Adam and assigned non-European peoples completely separate origins, such as to Cain, the sinning son of Adam (Graves, 2013). The theory of polygenism became a combined religious and sociopolitical claim that God intentionally created separate races and endowed an inborn hierarchy of race; while it was not a scientific theory, this idea was intentionally reinforced by pseudo-scientific tactics, such as measurements of cranial volume and lithographs contrasting racial types (Garrod, 2006). Several 19th century polygenist thinkers, such as Louis Agassiz, Samuel Morton, Josiah Nott, and George Robin Gliddon, were simultaneously pseudo-anthropologists and natural scientists who used anthropometry, skull measurements, lithographs, and photographs to craft and archive distorted depictions of a racialized other (Garrod, 2006). Agassiz and Morton were two key contributors to the development of scientific racism during this time.

Agassiz, an noted anti-Darwinist, set out to prove his belief in the fixity of race, or the idea that there existed distinct human races that could not be diluted over generations (Stepan, 2001). He led an anthropological study that used photography to archive people he assigned as either “pure” race or a “hybrid” combination of indivisible categories—specifically European, Indigenous, and African ancestry—within Brazil (Stepan, 2001, p.110). He used exploitative

tactics to take and archive primarily nude photos, attempting to delineate morphological differences that he could use to characterize racial "types" (Stepan, 2001). His discourse created dangerous concepts of people as "purebred" or otherwise mixed, arguing that "hybrid" individuals would always try to have children with a pure racial type to reverse this conceived hybridization (Stepan, 2001). Furthermore, Agassiz asserted that "hybrid" persons were prone to progressive degeneration across generations (Stepan, 2001). While Agassiz collected photographs, others recorded drawings and lithographs, which were also prone to curatorial bias. Samuel Morton's 1839 craniological atlas, *Crania Americana*, includes a lithographic comparison of two skulls, one from a Peruvian individual and one from a European. Morton attempted to visually delineate inherent differences between members of an "ancient race" and those of European modernity (Stepan, 2001, p.95). Lithographs were extensively used as an analytical tool, but could be distorted to create racist constructions (Stepan, 2001). Such analyses center the skull to evidence brain capacity and thus cognition or aptitude. His atlas makes flagrant, and heinous, scientific errors, such as the artistic contortion of the skulls as well as the reliance upon one skull, or one individual, to generalize to an entire community or race (Stepan, 2001). These focal points and methodological choices are not benign nor random; what distinguishes pseudo-scientific thinkers from scientific breakthrough is retroactivity rather than falsification.

While all realms of science must contend with confirmation bias, pseudo-science is inherently plagued by it, characterized by a premeditated selection of evidence and lack of controls or comparisons to justify what is already believed to be true. Conversely, scientific theory operates on the premise that no hypothesis or theory is ever confirmed but only corroborated by further evidence and always subject to potential falsification by new

observations. Even so, we also know that scientists who made valid contributions to their fields often additionally surmised racist applications of their ideas. Swedish scientist Carolus Linnaeus, who preceded the aforementioned polygenists and founded the taxonomic system that evolved into the biological classification system we use today, designated subtypes within *Homo sapiens*, including *H. sapiens europaeus*, *H. sapiens afer*, *H. sapiens asiaticus*, and *H. sapiens americanus* (Graves, 2013). He, like other 18th and 19th century naturalists, attached hierarchal meaning to these classifications, casting *H. sapiens europaeus* as "active and acute" and *H. sapiens afer* as "crafty, lazy, and careless" (as cited in Graves, 2013, p. 39). Fast forward to today's intra-species analogies to dog breed temperament and intelligence, and we can see an eerie preservation of continuity. There is a potency to the collective imagination, especially the scientific imagination, as people mistake scientific progress to be linear.

In contrast to the work of polygenistic and naturalist writers in Latin America, Charles Darwin was working on his evolutionary theorems from the mid to late 1800s by coalescing numerous observations and critical examinations that challenged contemporaneous explanations of naturally occurring species. While he may have originally believed in the fixity of species, Darwin overturned this idea by recording, during his travels and voyages, empirical evidence to the contrary (Costa, 2009). Most commonly cited today are his records of unique Galapagos species and the environmentally-tied advantages of finch, tortoise, and other animal variations within their Island niches; however, his thought process began to fundamentally diverge from ideas of fixity during his contemplations on domestic breeders. Costa (2009) describes Darwin's conceptual link between the breeding process and speciation:

> A key insight appears to be the breeders' assertion that "picking" over a period of time is sufficient to create new varieties—crossing, or hybridization, is unnecessary. Darwin's

> Malthusian insight comes about three-quarters of the way through the D Notebook, in passages dating to September 1838. (p.889)

This observation formed the basis for Darwin's next step, the idea that naturally occurring phenomena could produce the same nascence of variety as human intervention through this gradual "picking" of traits (Costa, 2009). Darwin eventually produced such a strong critique because, like Mendel, he provided extensive deductive evidence, researching not to prove but to discover. Nonetheless, like other naturalists of his time, Darwin looked up to Alexander von Humboldt as a highly influential figure (Costa, 2009; Garrod, 2006) as well as speaks of civilized and uncivilized races and links between skull capacity, "intellectual faculties" (p.145), and the "[proven]" (p.146) cranial differences between "Europeans", "Americans", "Australians", and "Asiatics" (Darwin, 1871, p.146). Humboldt's demeaning ideas of Indigenous peoples and the Tropical "other" subsisted in the popular consciousness, Darwin held onto foundationally racist conjectures, and fallacious social adaptations of Darwin's theory emerged.

The phrase "survival of the fittest" was crafted by Herbert Spencer, a sociologist who attributed the "growing social inequality in both England and the United States" (as cited in Garrod, 2006, p.56) to societal competition and differential human fitness. Overlapping concepts characterize Social Darwinism, an umbrella term encapsulating varying ideologies that apply Darwinian evolution to human sociology (Bowler, 2016). The basic idea argues that aspects of human nature are biologically ingrained and arose gradually over the course of past survivalist necessity. This concept has been applied to explain political systems, war, and race relations. In her chapter section discussing the "rising tide of white world supremacy", McWhorter (2009) argues that "science… portrayed racial conflict as an inevitable subplot in an evolutionary narrative" (p.142), and recurrent studies that described Indigenous peoples as "dying out" added

"more empirical weight to the claim that white world domination was imminent" (McWhorter, 2009, p.144). Evolution was employed to imagine race as a long-term process of human differentiation, supposing that groups of primordial humans broke away from a once unified species, sustained insulated civilizations, and independently evolved into contemporary racial groups with adapted, inborn characteristics (McWhorter, 2009). This idea reified environmental deterministic in a new context of heritability, supposing that the harsh, temperate landscapes of Europe enabled the European subgroup to become inherently civilized through evolutionary mechanisms (McWhorter, 2009).

Evolution proved to be a theory that was readily and falsely manipulated from the time of Darwin and through the 20th century. In 1962, Carelton Coon published *The Origin of Races*, arguing that different subgroups of people ascended the ladder of human evolution, and process of civilization, at different rates (as cited in Jackson, 2001). Using evolution to fabricate the innateness of biological difference and ingroup traits, scientific racism conveyed race as the product of a lengthy evolutionary process spanning far beyond the reach of scientific evidence and through the earliest beginnings of *Homo sapiens*—making the theoretical argument empirically inaccessible. And, this argument could be furthered by an imagined need to protect the evolutionary process from compromised bloodlines (McWhorter, 2009). And so, race relations, or racial conflict, were explained by Social Darwinists as evolutionarily codified over time, attempting to naturalize white supremacy, and its violence, as natural and inevitable. Social Darwinism lended to the development of "eugenics, pseudoscience, and psychometry" (Garrod, 2006, p.56) by Sir Francis Galton.With the advent of genetics in the 20th century, racialization adopted a new genetic language that cited invisible "factors of heredity" (Yudell, 2011, p.17) instead of the skull, skin, and other physical traits to establish race as inherent. These

rudimentary understandings of genetics further fettered race to biology and heredity and were solidified within eugenic ideology:

> This geneticization of race—the idea that racial differences in appearance and complex social behaviors can be understood as genetic distinctions between so-called racial groups—was shaped, in large part, by the eugenics movement. (Yudell, 2011, p.17)

The logic follows that if race, and the social factors it supposedly predisposes, is an inherent and inheritable genetic characteristic, then different races can be extricated from the human gene pool. Two coexisting branches stemmed from this movement, including the idea of positive eugenics, which promoted selective breeding between certain groups, and negative eugenics, which "denied the right to reproduce" (Yudell, 2011, p.18) from certain groups. Both were vehemently racist, ableist, and queerphobic at their core, representing strategies to create the idyllic version of humankind—the white supremacist version.

Eugenics is inseparable from fallacious concepts of heredity. In a perspective piece, Ryan (2015) describes eugenicists to believe that physical traits, and more specifically race, were indivisible from moral character, temperament, and aptitude, all inherited congruently unless diluted by dissimilar blood. In the United States, promoters of eugenics, such as R. W. Shufeldt (1915), disseminated the idea that the increasing social and political presence of Black communities should be America's primary concern, sensationalizing that the "aim and highest ambition" (p.124) of Black women was to have children with white men because "the superior intelligence coming from their white fathers, will command better positions… powerfully [furthering] the interests, political and otherwise" (p.124) of African Americans. Shufeldt attempted to fuel a racist hysteria against Black Americans while also reinforcing the idea that racially superior and inferior genetics existed. Charles Davenport tried to justify institutional

racism by fabricating the existence of a "constitutional, hereditary, genetic basis for the difference between the two races in mental tests" (Yudell, 2011, p.18), referring to Black and white Americans. Again, Davenport cites heredity and genetics as evidence for inherent racial disparity, and ideas like his formed a statutory basis for legalized eugenics. In the 20th century, the United States saw the advent of forcible sterilization laws, rooted in these false ideas of genetics and heredity, that amassed anywhere from 30,000 to 60,000 cases (Stern, 2016; Yudell, 2011). Many sterilizations were aimed at women deemed "feebleminded" or "insane", but coercive tactics were disproportionately applied to "Mexican American (Guitierrez 2008), Puerto Rican (Lopez 2008), Native American, and Black women (Solinger 2005)" (as cited in Shreffler et al., 2015, p.3) during the early to mid 1900s. Eugenicists also helped fashion segregation laws in the United States throughout the early 1900s. Among the most prominent voices of resistance were African American scholars and activists Kelly Miller, the Dean of Howard University, and W. E. B. Du Bois, who fought for humanity, scourged the flagrant scientific fallacies of eugenicists and challenged biological concepts of race as, instead, socially constructed (Yudell, 2011). Indeed, understanding race as scientifically imagined, but with social consequences, is the key to dismantling ideologies of biological race. Ironically, the conceptual products of racialization that were reinforced by scientific racism during the 18th through 20th centuries actually serve as prime examples of how race is non-contiguous.

The social construction of the American race concept is apparent in contrasts between peoples within the US as well as in comparison to other countries, calling upon a common thread of blood, purity, heredity, and ancestry. Over the course of American history, Euro-American colonizers racialized Native American tribal membership, and antimiscegenation laws regulating racial distinctions remained intact until around 1968 (Tallbear, 2013). The 19th century saw

political efforts to "detribalize" Native Americans through assimilation projects requiring boarding-school education, prohibiting religious practices, and dismantling communal living infrastructures, all backed by the pseudo-scientific idea that Native American blood could be "diluted" over multiple generations (Tallbear, 2013, p.47). This notion of dilution perhaps harkens back to the ideas of Christopher Columbus and his contemporaries during preliminary encounters with the Americas:

> Columbus's descriptions of weak innocents and fierce cannibals established a dichotomy that framed most European characterizations of the Native people of the Americas for the next five centuries and more. Despite depictions that distinguished sharply between Europeans and misnamed 'Indians' at the outset of colonization, many Europeans believed the latter could be transformed. Three "sauage men" from northeastern North America arrived in England in the 1490s "in their demeanour like to bruite beastes,' Robert Fabian related, but after two years he "coulde not discerne [them] from Englishmen." (as cited in Harvey, 2016, p. 3)

Fabian's racist statements and caricatures reflect the later notion that Native American identity could disappear, or be assimilated, into European whiteness over the course of subsequent generations; the social and political factors surrounding this theory will be discussed more deeply in Chapter 6. In the 20th century, the US also conjured the one-drop rule for African Americans, whereby "racial distinctions were based strictly on skin colour" that categorized "anyone visibly 'non-white' as 'black'" (Stepan, 2001, p.94). This rule was born in the South and eventually pervaded nationally, classifying Blackness based on the presence of a single African ancestor in one's lineage (Davis, 1991). The rule was a manifestation of racist Jim Crow anxieties that further barricaded African Americans from resource access and societal

participation in spite of pushes for the assimilation of other groups (Davis, 1991). The one-drop rule contrasts the dilution theory applied to Native Americans, evidencing that the social malleability of racialization depends upon interests of the hegemonic group. A key difference between the two race concepts centers tribal land sovereignty within colonial America, as a deemed loss of Indigenous identity could be leveraged to diminish land ownership (Steers-Mccrum, 2018). Thus, race could be constructed differently depending on place and as shaped by history, naturalist discourse, colonialism, and sociopolitical interests. Brazilian race structure differed from the "white/non-white dichotomy" of the United States and instead held quandrants of varying racial intersections and component proprotions (Stepan, 2001, p.105). Numerous terms emerged to describe different peoples, such as "pardon, cafuzos, mestiços, and morenos" (Stepan, 2001, p.105). Specificity of language often reflects the social legitimacy as well as societal relevance of termed identities, and that reigns true for the race concept of Brazil as well as other Latin American countries post-independence. In a study on the "Ideology Of White Racial Supremacy: Colonization And De-colonization Processes" in Brazil, Nogueira (2013) explains racism as a more nuanced color gradient whereby "the darker the skin color, the more one suffers discrimination" (p.25). Moreover, whiteness is less tied to ancestry and more strictly ascribed to phenotype and social status in Brazil (Nogueira, 2013). The process of racialization, and histories of racism and colonization, manifested differently in both countries, enabling the social and interactional nature of race to be evidenced through comparison.

Scientific racism culminated on a global scale when negative eugenics manifested as genocide by the Nazi regime, feuling a war for so called 'Aryan' (white) supremacy (Yudell, 2011). Eugenics and genocide clearly visualized the evils of scientific racism (queer, disabled, Polish, Catholic, Serbian, Soviet, and all other persons targted by the Nazi regime should not be

forgetten here) at the cost of immense suffering and innumerable lost lives. Following World War II, efforts to critically analyze and decisively sever notions of biological race emerged. New developments in genetics, such as the discovery of the structure of DNA, study of multifactorial and epistatic genes, the mapping of the human genome, and the nascence of epigenetics undermined simplistic and unsound ideas of human heredity. In 1954, the *Brown v. the Board of Education* US Supreme Court decision cited key evidence in an influential text by Gunnar Myrdal, an economist from Sweden, who attested to the "great variability of traits among individuals in every population group… and the considerable amount of overlapping between all existing groups" (Yudell, 2011 p.21). The revelation that human genomes showed 99% similarity across all communities prompted non-scientific world leaders, such as Bill Clinton (*Human Genome Sequencing*, 2000), to announce that "all human beings, regardless of race, are more than 99.9 percent the same". Genetics appeared to now be unraveling ideologies of difference entrenched in scientific racism—yet, it took the perceived final authority of science for people to reimagine race.

Again, we pivoted. We discerned a minute, but detectable, window of allele frequencies that were unique to specific cohorts, and, as it follows, certain ancestral lineages. We honed in once again on the science—the genetics—of difference. The process of racialization has delimited people by physical appearance and ancestry, emphasizing the collective, but fabricated, inheritance of intellectual, moral, and otherwise immutable traits. Such ideas fundamentally center heritability, paralleling in concept the immutable nature of alleles and the vocabulary genetics at large (albeit a grossly fallacious version of genetics). Perhaps we should come to understand the history of race in science not as a subsection of the race concept, or a byproduct of Eurocentric scientific histories, but as fundamental to creating much of the scaffolding of

racialization itself. Biological race was not only purported by pseudo-scientists (i.e. Agassiz) but also reinforced by those who produced work deemed legitimate and preliminary to currently accepted theories but nonetheless injected heinously racist ideas (i.e. Linnaeus). While authentic displays of the scientific method will always fail to delineate biological race, the field of human population genetics is running into complex problems of racial conflations because, amidst searching for the tie between allele frequency clusters and probabilistic inheritance, the field is having trouble emerging from a racialized lens. This difficulty is especially apparent in attempts to construct scientifically relevant organizations and cross-comparisons of genetic ancestry.

Today we see a resurgence in genetic discourse surrounding perhaps not explicitly biological race but proxies of genetic ancestry, continental origin, and admixture. Steadily advancing genotyping and chip technology has enabled geneticists to investigate the minute window of potential interpopulation differences that were previously inaccessible. Furthermore, there is a pressure within disease-susceptibility research to control for genetic substructure that may have arisen overtime. As we have seen in Chapter 1, geneticists uphold an intricate, logical basis for the study of genetic ancestry. Human cohorts have likely at different points in time and different places experienced random effects of genetic drift or selective effects of regional pressures[13], but the descendent imprints of these effects do not fit neatly within any existing concept of either race or ethnicity. Geographical origin might provide a comparatively better predictive quality but still bears technical errors and can easily be conflated with or erroneously used in place of race. But, if ancestral substructure does exist, we are likely asking the wrong questions - looking from within the wrong framework. The ways in which we are talking about human populations and genetic admixture are echoing antecedent ideas of racial types, subtypes,

[13] Selective pressures could refer to instances such as the relationship between sickle-cell anemia heterozygosity and Malaria, where the codominant expression of sickle-cell anemia confers immunity from Malaria.

and proportions. We are not effectively capturing new, quantitative concepts of allele frequency differences in terms of how they could arise over space and time but rather focusing on a static image of ancestral vestiges in modern day differences. We are not talking about population genetics in a way that decisively eradicates race from scientific discourse. The next chapter will take a comparative approach to the scientific rhetoric of modern genomics. Drawing upon rhetoric of the past, I will investigate how genetics discourse today sustains the continuity of or re-conceptualizes the race concept, both explicitly and implicitly, in primary literature, patent applications, and scientific addresses from the late 20th century to present day.

CHAPTER THREE

Continuity

Critical Examinations of Present Genetics Discourse

Past scientific imaginations of the race concept continue to inform how the natural sciences discuss, or selectively omit, race, ethnicity, and ancestry. From the late 1900s and through modern day, primary literature, database projects, and institutional statements have addressed the intersection of genetics and race, among other sociocultural identities. Whether aiming to combat scientific racism or define a new kind of intergroup difference, modern genetics contends with a racialized framework reinforced by the past several hundred years. Most geneticists try to decisively separate their research from race, but lines of continuity have resurfaced, specifically amidst the language surrounding human populations. Within their primary literature, researchers grapple from within the confines of an accustomed scientific vocabulary bred from past ways of organizing difference and taxonomizing race. The International HapMap Consortium, a collection of scientists and funding agencies collaborating on one of the largest genetic database projects to date, made a statement in 2004 discrediting specious interpretations or uses of their project data:

> ...the lack of precise population definitions and ancestral geographical self-assignment is not problematic for developing the HapMap because exact demarcations are not necessary for the way that the HapMap will be used, and the Project does not aim to define populations. No claims are made about the genetic 'purity' of the sample sets or the populations of which donors are members; such claims would be scientifically

> spurious, as human populations are the products of countless social, historical and demographic processes, and therefore can never be sharply defined. (box 2, p.469)

The International HapMap Consortium takes an important position that denies genetic "purity" and emphasizes the complexity of heritage. But, the researchers that use genetic databases must be able to affirm that they have gathered samples representative of different populations for comparative analysis of interpopulation allele frequencies to make sense. Likewise, projects such as HapMap still have to make selective decisions about who they will study and create working definitions of human populations. In fact, the Coriell Institute for Medical Research outlined "Guidelines for Referring to Populations" that stated:

> The way that a population is named in studies of genetic variation, such as in the HapMap or 1000 Genomes Projects, has important ramifications scientifically, culturally, and ethically. From a scientific standpoint, precision in describing the population from which the samples were collected is an essential component of sound study design; the source of the data must be accurately described in order for the data to be interpreted correctly. From a cultural standpoint, precision in labeling reflects respect for the local norms of the communities that agreed to participate in the research, and an acknowledgement that populations in one part of the world are not all the same. (*Guidelines for Referring to Populations*, n.d., para. 2)

The guidelines, which surround the use of the Coriell Institute's NHGRI Sample Repository that draws from HapMap and 1000 Genomes populations, go on to emphasize the ethical importance of precision in preventing under- and over-generalization of results as well as promoting language consistency across studies (*Guidelines for Referring to Populations*, n.d.). These guidelines paradoxically contrast the International HapMap Consortium's viewpoint, the former

emphasizing the need for precision while the latter conceives imprecision as relatively unproblematic.

In their study on the "Application of Genome-Wide Single Nucleotide Polymorphism Typing: Simple Association and Beyond", Gibbs and Singleton (2006) review the genome-wide SNP assay, a robust and efficient technique useful for association studies and made possible by the accumulation of SNPs for the HapMap project (p.1511). Gibbes and Singleton (2006) reference the purpose of the HapMap project as determining how "SNP tagging approaches may vary from population to population" by generating in their first phase "four discrete populations: Yoruba from Ibadan, Nigeria (YRI); Japanese in Tokyo, Japan (JPT); Han Chinese in Beijing, China (CHB); and Utah, United States, residents with ancestry from northern and western Europe (CEU)" (p.1511). Even within a direct application of the HapMap project, human populations are being characterized as "discrete"; furthermore, Gibbs and Singleton (2006) elucidate that the first phase HapMap populations were distinguished observationally without prior knowledge of genomic profiles that existed within the four communities. The study fails to demarcate human groups as preceded by numerous "social, historical and demographic processes" (2004, box 2, p.469) the International HapMap Consortium referenced, which would help readers understand that attempts to identify study populations are based on extrinsic predictions of comparative genetic difference. The purpose of a database is to supply archived sample DNA to genetic studies, but we can see that one degree of separation already muddles the warning message of the International HapMap Consortium. The remainder of this chapter will look at a couple examples of primary literature, database projects, and scientific addresses as case studies of scientific rhetoric within the field. The scope of this chapter is limited to a small number of examples, as its primary purpose is to initiate conversations about how current

rhetoric reifies past racial typologies through new sociocultural conflations rather than speak comprehensively on the study designs or intentionality of all contemporary genetic research.

With an increased focus on genetic populations, researchers in the field launched their own diversity projects. Earlier versions were not as tied to biomedical or other forms of applied research but took a broad interest in explorations of human history. On September 21st, 1994, Luca Cavalli-Sforza, then a professor of genetics at Stanford University, made an address to UNESCO regarding the Human Genome Diversity Project. At the time, the project self-defined as an anthropological endeavor to study the "genetic richness of the entire human species" (Cavalli-Sforza, 1994, p.1) by accumulating globally representative genetic samples, but it would eventually evolve into genomic analysis of human variation and a cell line panel including over 1,000 individuals from 52 populations worldwide by the 21st century (Rosenberg et al., 2003). Within the address, Cavalli-Sforza outlines its main goals and methods, citing one of the primary roles as "[combating] the scourge of racism" (Cavalli-Sforza, 1994, p.1). He believed that studying human population genetics was one of the best ways to prove the wrongness of racism (Cavalli-Sforza, 1994). Scientific racism entrenched social concepts of the racialized other in fabricated scientific "fact"; in other words, it was a social manipulation that assumed a scientific voice. But Cavalli-Sforza, among others, believed that science was the primary, and perhaps only, tool to reverse this trajectory, the only means of summoning counteracting proof. Concepts like these, though well-intentioned, miss gaps filled by a sociological perspective, which understands the human role in crafting race and power systems.

While late 20th century geneticists like Cavalli-Sforza discredited past fallacies of biological race, they believed there was information to be drawn from scientifically-relevant populations. Cavalli-Sforza (1994) notes the central but somewhat elusive genetic basis of

human populations: “in general, human populations differ, with respect to a genetic marker, only in the relative frequencies of the different forms” (p.4). He explains that any given individual in any “population” could inherit any of the possible forms of a gene, but groups of individuals can be characterized by the collective proportions of each form. One group could theoretically differ from another group or the global population at large by the frequency of one or multiple gene forms. This definition sheds light on the quantitative and probabilistic nature of genetics, as groups could technically be defined arbitrarily by numerous different metrics, and in-group allele frequencies are subject to constant change with each subsequent generation. We see the distinctions of Cavalli-Sforza’s definition begin to dissolve when the HGDP translated precepts into practice. The search for groups where allele frequencies have been structured by common ancestry led the HGDP back to a lens of coloniality, looking to the same places where previous naturalists perceived difference.

The HGDP prioritized groups that were “anthropologically unique” (Cavalli-Sforza, 1994, p.7) and practiced languages or cultures that made them distinct from most or every other population. Caballi-Sforza thought that studying these groups would ultimately yield benefits for both the community members themselves and all of humanity by learning from preserved genetic “uniqueness” (Cavalli-Sforza, 1994). This paternalistic intent translated into the seeking out of Indigenous tribes, clans, and communities worldwide, raising important ethical and social concerns, specifically intersecting with boundaries of Indigenous sovereignty and identity (Barker, 2004; Dukepoo, 1999). Cavalli-Sforza understood “isolated” populations, perceived as culturally, linguistically, and geographically bound, to hold more information regarding human evolutionary history; further, he worried that convergence, or mixing, of these communities with others would curtail the information they could provide in the now (Barker, 2004). In her study

on the cultural politics of identification in the Human Genome Project, Joanne Barker (2004) notes the continuity in the prioritization of populations:

> Regardless of the "good intents" of those involved with the HGDP, the lack of reflection on their discourses disavowed the very realities of colonialism and racism that produced the situations that put these various populations on the HGDP's list in the first place. (p.586)

Barker denotes here that their perception as a distinctive other has not dissipated, specifically amidst the HGDP prioritization schema. The "list" she mentions refers to the potential existence of an actual list of several hundred target groups characterized by their likelihood to "disappear", whether due to natural disaster, war, or merging and dispersal (Barker, 2004). Since early contacts between colonizers and colonized land, Indigenous people have been objectified to scrutiny, scientific dehumanization, violence, and displacement, and Indigenous tribes continue to face power imbalances and injustices in science today. Not long before Cavalli-Sforza's address, Arizona State University researchers studying Type II diabetes asked members of the Havasupai Tribe for DNA samples. Tribal members agreed with the understanding that their donations would facilitate diabetes research, but "in 2003, it was discovered that the samples were used for studies on schizophrenia, history of migration, and inbreeding" (Lee, p.147, 2015) without releasing these intentions explicitly in the consent form or gaining permission from the tribe. A lawsuit was filed (*Havasupai Tribe v. Arizona Board of Regents*) and the Havasupai Tribe obtained compensation as well as the original DNA samples (Lee, 2015). The experience of the Havasupai Tribe exemplifies the continuity between colonial exploitation and modern genetic research, elucidating the dangers posed by the targeted scrutiny of the HGDP.

Controversy still surrounds the HGDP, and many critics today understand it as a failed attempt to archive culturally, linguistically, and otherwise socially diverse DNA banks that generated ethical and consent-related concerns (Foster & Sharp, 2002). Today, genomics has entered an era of next generation sequencing (NGS), characterized by unprecedented analytical breadth and efficiency (Pardo-Seco et al., 2014). Marked technological advancements, such as "chip-based genotyping" (Saeb, 2016, p.1), and reduced costs have enabled significant database growth for projects universally, including the HGDP. NGS enabled the transformation of the HGDP into the HGDP-CEPH human genome diversity cell line panel, which has been used in papers from the 2000s through the present to integrate human demographic histories within clinical and forensic genetics (Holsinger & Weir, 2009; Pereira et al., 2019; Shi et al., 2009). Several concurring database projects exist, the HGDP representing an early version emerging alongside first full sequencing of the human genome.

Large-scale genetic databases, which echo similar purposes of increasing sample diversity, have normalized discourses surrounding human populations and sampling methods as well as supported database compilation as a central interest of modern genomics. Such projects include the HapMap project and the 1000 Genomes Project as well as fine-scale studies forming databases within countries and among regional ethnic groups. Among other databases, NGS has catalyzed significant developmental strides in the HapMap project, purposed with accumulating vast population-specific SNP data to facilitate association studies; HapMap has entered its third phase, seeing the addition of 1 million SNPs to both registered and new population classifications, and has mapped a vast span of rare and common gene variants to enable the continued study of population differentiation (Altshuler & Donnelly, 2005; Elhaik, 2012; Saeb & Al-Naqeb, 2016). Applications include those by Elhaik et al. (2012), who re-examined Fst

distributions using 3 million SNPs from eight populations and 1 million SNPs from globally representative population compared to the upper limit of 40,000 SNPs that studies throughout the 2000s used. Pardo-Seco et al. (2014) used the entire HapMap database to proxy as a global sample their inter-continental analysis. These kinds of database advancements are the gateway toward more expansive study design, as studies can simply download population datasets (as they have been classified within the particular database) directly from project web sites. Thus, genetic databases have a central role in framing the genetic ancestry concept. Sandra Soo-Jin Lee (2015) explores the social and political exercise of termed biobanks:

> In addition to being physical repositories, biobanks are unique social artifacts that concretize assumptions about population boundaries to organize and sort human samples. Informed by sociohistorical taxonomies of how to distinguish human groups, policies on categorizing biological materials stored in biobanks reveal the shifting and contingent meanings around genetic differences and their significance for concepts of ancestry, ethnicity, and race. (p.144)

Genetic databases, or biobanks, form the initial "taxonomies of 'molecular difference'" (Lee, 2015, p.144) that shape how studies perform association analysis. Database groupings are often created before the DNA samples belonging to each group are collected, allowing database projects to actively control who is prioritized for study and who, at large, is predicted to show genetic differences within the minute window of interpopulation variation. They perform a racializing refraction of how we interpret the very term "interpopulation". A major fallacy lies in defining individual membership to groups based on genomic profiles, as any semblance of interpopulation difference can only be examined at the group level. Nonetheless, molecular taxonomies have led to a massive focus on ancestral inference. These databases serve as an

integral scaffolding for the study of genetic ancestry and the standardization of rhetoric surrounding human populations.

After HGDP evolved into the HGDP-CEPH cell line panel, contemporary uses and applications of the Human Genome Diversity Project, among other modern databases, reimagined past concepts of discrete genetic elements in another branching line of continuity. While Cavalli-Sforza (1994) discredited notions of genetic "purity", explaining the near impossibility of generating any such thing even with plants and animals in the laboratory, his fears of a fleeting opportunity to study isolated populations reimagined concepts of purity and intercrossing but in terms of tapping older, undispersed, or undiluted ancestral lineages. Reification of these concepts are majorilly apparent in the ways modern geneticists try to categorize people. In a recent study entitled "Inference Of Human Continental Origin And Admixture Proportions Using A Highly Discriminative Ancestry Informative 41-SNP Panel", Nievergelt et al. (2013) ran validity tests on an AIM panel, drawing samples from both HGDP-CEPH and HapMap reference populations. The researchers wrestle with keeping concepts of human population distinct from old notions of typology and mixing. When describing genetic maps of "Native American populations" in the Americas, Nievergelt et al. (2013) note that "admixed Muscogee and HapMap Mexicans" did not cluster as expected and "showed a strong European component" (p.7). Here, the researchers try to qualify geographic origin through genetic ancestral inference, but describing a "European component" raises ambiguity. What does "component" mean? It implies that one's heritage can be numericized as parts of a whole, that convergence between ancestries yields a fractioned inheritance, and that European ancestry can be captured as strong or weak within a genealogy.

The answers to these questions are neither self-evident nor explicit within the body of the article. Nievergelt et al. (2013) present their findings on "admixed Muscogee and HapMap Mexicans" as an noteworthy exception to the predicted genetic clustering of Native American tribes. A hundred years prior, H.M. Tomlinson, an Englishman who explored the Brazilian Amazon in 1909, recorded his thoughts on the individuals he encountered while traveling in the providence of Pará, expressing confusion as "one used to the features of a race of pure blood" (as cited in Wood & Chasteen, 2009, p.147). Rooted in ignorance, Tomlinson is perplexed by people who challenged his internalized constructs of race. He invokes ideas of pure, indivisible races, recording (using different terms) that he saw Portuguese, Afro-Brazilian, and Inginenous-Brazilian components "but rarely a true type of one" (as cited in Wood & Chasteen, 2009, p.147), as if race could be reduced as predeterministic ingredients. He tries to physically characterize which features come from each "pure" race, observationally cataloging and dissecting real people. In doing so, Tomlinson believes that he could discern those in whom "[B]lack was the predominant factor" (as cited in Wood & Chasteen, 2009, p.147) and those in whom it was not. The terminology of "predominant factor" is not too distinctive from "strong European component". Both make the reader visualize ancestry as an assemblage of distinct or blended components. Thinking back to the descriptors used by Nievergelt et al., we can also pose the following question: how many ancestors does it take for a "component" to be conceived as a whole, or as the dominant identity within which other ancestries are "component"? Nievergelt et al. (2013) clearly contend with the fundamental quandary of devising "more objective and accurate methods of defining homogenous populations" for scientific study (p.2). They recognize a need for improvement, but this statement begs the question of how to define homogeneity in the context of humans and the scale by which relative differentiation or relatedness constitutes

homogeneity. What is the threshold of difference, or what proportion of loci, permits a geneticist to subdivide groups, and what is the threshold of relatedness if most genetic variation occurs across all human individuals? Furthermore, the idea of a "European component" safeguards geneticists from explaining the processes of European colonialism and sexual exploitation that lent to describing a "European component" in the genealogy of Indigenous families. Identifying where dated language resurfaces in present literature not only prods at unanswered questions but also helps unveil the elusive framework that continues to guide studies of human difference.

Reuse of outdated scientific archetypes primes readers to conceptualize new findings through the same racialized lens. In 2011, Dorothy Roberts accentuated this contention, drawing focus to the advancing technical capabilities of genomics in a Human Genome Project postmortem, indicating that "biological theories of race…" are experiencing a revival through "cutting-edge genomic research… [modernizing] old racial typologies" (as cited in Nelson, 2016, p.13). In their 2013 article, Batai and Kittles discuss whether genetic ancestry, as it is defined and detected using current research methods, provides practical insight into understanding health disparities. As the founder of African Ancestry tests, Ph.D in human genetics, and lecturer on Afrocentric genetic ancestry, Dr. Rick Kittles believes there is promise to genetic ancestry, but he and Batai are critical about its implications and made an important note that modern geneticists and researchers are still referencing racial categorizations from scientific studies on race that predate modern genetics and primarily distinguished physical characteristics (skin color, facial contour, and skull parameters) (Batai & Kittles, 2013). Roberts, Batai and Kittles are accurate in their reference to the use of outdated constructs, as racial typologies have resurfaced both subtly and overtly. In a 1982 article on the "Evolution of Human

Race at the Gene Level", Masatoshi Nei uses protein and blood group loci to try to qualify gross genetic relatedness using past, rudimentary racial types:

> Genetic distance data indicate that Caucasoid and Mongoloid are somewhat closer to each other than to Negroid. Analysis of restriction site data for mitochondrial DNA also shows the same genetic relationship. It seems that the Negroid and the Caucasoid-Mongoloid groups diverged about 110,000 $\pm$ 34,000 years ago, whereas Caucasoid and Mongoloid diverged about 41,000 $\pm$ 15,000 years ago. (p.167)

These terms invoke colonial era archetypes of human origin; it seems that, amidst the advent of molecular genetics, they were unwittingly relied upon once again, this time assumed to be genetically-relevant categories. But, even as recent as 2005, we see an overt reinsertion of racial typologies to characterize the "ethnic origins" and multiple identities of India for the Indian Genome Variation Database (IGVdb) (2005):

> Indian population can be, to a large extent, substructured on the basis of their ethnic origin as well as linguistic lineages. All the four major morphological types—Caucasoid, Mongoloid, Australoid and Negrito are present in the Indian population (Malhotra, 1978). The "Caucasoid" and "Mongoloid" populations are mainly concentrated in the north and northeastern parts of the country. The "Australoids" are mostly confined to the central, western and southern India, while the "Negritos" are restricted only to the Andaman Islands. (p.2)

Both Nei and The Indian Genome Variation Consortium refer to the same subset, generally translating into Europe, Africa, Australia, and Asia; Nei (1982) mentions "Australoid" later in the text. Purported racial typologies have shifted in number and scope, but seem to always center around 3-5 main groups (interestingly, the first three ancestries HapMap intended to study were

African, European, and Asian ancestries). The original uses of common typological terms date back to 1795, when Johann Blumenbach published his third addition of *De generis humani varietate nativa*; this edition delineated "five generic varieties" comprising "Caucasians", "Mongolians", "Ethiopoians", "Americans", and "Malays", whose physicalities he extensively detailed according to the intra-group anatomical similarities he perceived (Bhopal, 2007; Takezawa, 2012). Blumenbach tried to denote them as varieties without static bounds, but he still construed racialized meaning from perceived physical differences[14], and scientific racism later perpetuated these types within its discourses (Takezawa, 2012). For example, we have seen Carolus Linneaus echo a similar schema in his taxonomic conjecture of four foundational subgroups within *Homo sapiens*, both linked to geography (Garrod, 2006) and entrenched in an "ideology of race that is used to explain, predict, and ultimately control social behavior" (Lee, 2015, p.146). And, the convergence of novel genetic methods with vestiges of racial typology becomes especially concerning in light of contemporary arguments, such as Nicholas Wade's *A Troublesome Inheritance*, a book published in 2014 that harkens back to discourses of environmental determinism; Wade defines a set of biologically distinct human races that are tied together by distinguishable social behaviors and temperaments, assigning Western civilization as superior due to the entrepreneurial nature of the Caucasian race (as cited in Nelson, 2016, p. 14). Continuing to use these terms in genetic literature reinstates their intellectual authority, homogenizing broad continental regions as distinct racial groups.

[14] Blumenbach's (1865) works disseminated the term "Caucasian" (p.249) as representative of "the most beautiful race of men" (p.249) due to cranial structure and whiteness, which he believed to be the primordial color of humankinds since it is easy for white to "degenerate into brown" (p. 249). His term "Caucasian" is geographically rooted in the individuals he observed at Mount Caucasus between Europe and Asia, yet he extrapolates such instances as a fundamental human "variety", conveying how easily geographic origin can be racialized.

In more subtle cases, certain studies have assumed a central focus on five to six "main" continental regions as genetically-relevant delineations. Along with other geneticists, Rosenberg et al. (2002) concluded that differences between individuals within a given population accounts for the majority of total human genetic variation. One of the earlier studies of population structure, this article detects six major clusters of genetic similarities, with five "[corresponding] largely to major geographic regions" (Rosenberg et al., 2002, p.2382). The study ran statistical analyses of 4199 alleles and visualized either across all or in only one of the following regions: Africa, Europe, the Middle East, Central/South Asia, East Asia, Oceania, and America (Rosenberg et al., 2002). Regional geography might be a pertinent comparative lens if we think about spatial differentiation and the random structuring effects of genetic drift; however, researchers have a significant hand in curating as well as making sense of the selected geographic boundaries, which interestingly often follow similar trends as the antecedent typological locales. The sixth genetic cluster that appeared at K=6 was perceived as an outlier because it did not conform to any of the expected "major" geographical regions:

> However, the next cluster at K=6 did not match a major region but consisted largely of individuals of the isolated Kalash group, who speak an Indo-European language and live in northwest Pakistan. (Rosenberg et al., 2002, p. 2382)

The idea of major, minor, or outlier geographical regions and ethnic groups is an ingrained organizational construct, built into theories of human migration across places before they were given names. We see another equivocal definition emerge from the study by Nievergelt et al. (2013), which references "outliers of minority ancestries" (p.14) as those existing outside of the designated main continental groups. They do not further define "minority ancestry", and it is unknown whether this refers to specific locations, ethnic or multi-ethnic identities, or "outlier"

clusters that are only apparent after "[grouping] subjects into continental clusters using a majority criterion" (Nievergelt, 2013, p.14) and adjunct statistical methods. While the science is vastly different and increasingly more complex, there is a critical aspect of modern genomics that is reminiscent of past efforts to observe and produce taxonomic schemas (Hauskeller et al., 2013). While helpful for visualizing complex theories on the history of humankind, today's commonplace ideas of major geographic groupings can also be constrictive, framing where we look or expect to see patterns and differences arise as well as how we construct comparative analysis.

As the study of human population genetics progressed, researchers had to develop methods for defining who constitutes a given human population, producing different versions. In their published patent application, "Computations and Methods for Inferring Ancestry", Mark Shriver, a biological anthropologist, and Tony Frudakis, a molecular biologist, (2004) promote the use of "biogeographical tests" in determining the relative proportion of ancestry in an individual. Although most versions attempt to sever controversy from genetics, Shriver and Frudakis (2004) conflate race within the conceptualization of biogeographical ancestry. Shriver and Frudakis (2004) precipitously assert the viability of inferring ancestry from physical traits, disease predispositions, and drug responsiveness as well as decisively interpolate race within its definition:

> On a basic level, human population structure can be represented in terms of BioGeographical Ancestry (BGA), which is the heritable component of "race" or heritage, and which is relevant on any scale of resolution. For example, on a crude level, BGA can be determined for 2 groups (e.g., European vs. others); or on a fine level, e.g., it can refer to "race" in terms of 4 groups. Such as IndoEuropeans, East Asians, Sub-

> Saharan African and Native American; or on a finer level, e.g., it can refer to ethnicity within the European group (for example, Mediterranean or Scandinavian).... (p.1)

Frudakis and Shriver communicate an acute, determinist testament to the heritability of race. The language utilized to describe human population structure appears dubiously ethnocentric, describing groups in terms of "European vs. others" or grossly clumping regional ancestry into broad categories of "IndoEuropeans, East Asians, Sub-Saharan African and Native Americans" (Frudakis, 2004, p.1). In describing fine-scale population structure, Frudakis and Shriver (2004) define populations by intragroup ethnicity rather than race, sustaining an unsound semantic leap between two sociocultural concepts as scientifically equivalent.

In one of their cited biogeographical tests, an experimental sample of participants self-reported their ancestry by classifying their parents and grandparents as the following: "'African', 'American Indian', 'Asian', 'Caucasian', 'Hispanic' or 'Other'" (Frudakis, 2004, p.39). Again, we see conflations of different categorization schema outlined by the test, and a number of incongruencies and connotations are apparent. "Hispanic" here could either refer to those of Spanish descent, those in any Spanish-speaking countries, or anyone identifying as Latinx. Today, some argue that the use of the word "Hispanic" inflates Spanish heritage, or Spain's impact as a colonizer, and dilutes the many racial and ethnic identities comprising Latin America and Latinx communities; however, use of the terms Hispanic, Latino/a, and Latinx[15] remain debated within communities and between individuals (Simón, 2018). Caucasian again references old racial typologies of Blumenbach and his contemporaries. "American Indian" invokes misuse of the term "Indian" by colonizers in the Americas and serves as the only Indigenous designation specified within the group, erasing innumerable Indigenous peoples worldwide. Finally, they

[15] Latinx is a gender neutral iteration that broadens inclusivity beyond binary constructs of gender.

create the categorical “Other”, which literally others those who do not identify with the prescribed categories. It is as if these categories are frozen within a colonial timeline, assuming the circumscription of these groups within their respective communities and a lack of immigration, integration, or space for multiple identities. Although Frudakis and Shriver recommend biogeographical ancestry (BGA) as a more scientifically objective classification schematic in place of race or ethnicity, their attempt to describe the “biological” components of race in the guise of BGA veils the sociopolitical realities of identity and exposes their ethnocentric lens. These initiatives, and the messages they disseminated, are components to the intellectual foundation of genetic ancestry in modern genomics. Frudakis and Shriver’s BGA test raises important issues surrounding the reliability of genetic classification schema, which will be explored further in Chapter 5.

Biogeographical ancestry is a new term for an old concept, namely biologically discrete elements of race. A more recent study by Pardo-Seco et al. (2014) discredit genetic support for race but continue to adapt biogeographical ancestry (BGA) as a more objective metric than other descriptors of human populations. Aside from the more obvious noise of analytical techniques, visualization software, high-throughput sequencing ability, Pardo-Seco et al. reference Pääbo (2003), who distinguishes how our understanding of genetics at the level of the genome, individual, and population diverges from past simplistic views of genetic elements and biological human groups:

> To understand what make us unique, both as individuals and as a species, we need to consider the genome as a mosaic of discrete segments, each with its own unique history and relatedness to different contemporary and ancestral individuals. (p.409)

Here, Pääbo (2003) provides an interesting map to understanding the genome as a complex process and product of inheritance, reminding us that both long-term ancestral heritage and close relatedness contributes to the multifaceted genetics of each individual. Furthermore, this understanding focuses on the genome rather than the individual, circumventing the idea that an individual is holistically defined by some sort of homogenous genome. But, when genetic ancestry translates into inclusion criteria, such as Frudakis and Shriver's BGA test, the standardization and consistency required for empirical study frequently gives rise to the modern versions of "old racial typologies", as previously described by Roberts. Rather than understanding allele frequency differences as impermanent and, at their core, unintelligible using sociocultural constructs, we get caught in the weeds of trying to attribute genomic segments to modern-day signifiers, shaped by racialization, colonialism and state-building.

The study of genetic ancestry has reopened conversations on racial and ethnic identity in science. Replacing "people" with "population" and "individual" with "sample" has become commonplace, and as a result, makes it easier to overlook the humanness of those implicated in genetic studies. In an interesting chronological quandary, we try to craft a working definition of "population" to enable interpopulation analysis, as Barker (2004) argues that populations are not, in fact, natural formations:

> The truth is that populations are not naturally occurring. In part, they were produced in the context of state efforts to classify and track various classes of people, including criminals, the poor, the diseased, et cetera, in the development and administration of state services and moneys (Hall et al. 1978, Foucault 1979). Within the emergence of the gene and genome as specific fields of study in biology and anthropology, populations were put to work in efforts to deracialize human genetics. They were called upon specifically

> during the incredible scrutiny over the role of eugenics in state programmes of genocide, mutilation and sterilization that followed the disclosure of the inhumane atrocities committed by Nazi scientists during World War II (Haraway, 1989; Hayden, 1998; Peters, 1998). (p.577)

Barker hones in on continuity, positioning the attempted deracialization of human populations in genetics as a reaction to the culminating atrocities of eugenics. The construct of populations creates a surrogate to study group difference as a scientific object insulated from historical context, politics of identification, and social realities. The quantitative allele frequency clusters that geneticists are trying to discern via populations likely exist in numerous, intricate, and transient ways but not in the static and racialized manners they are being conceptualized, which is especially evident in the discourse surrounding admixture. In a simulated scenario, Pardo-Seco et al. (2014) describe the conditions for admixture, using the term "hybrid AA-genomes" to represent admixed genomic profiles. Although hybrid specifically refers to DNA segments rather than whole persons (building off the idea of genomes as mosaics of pieces with independent histories), the choice of the term hybrid indirectly invokes Agassiz's ideology of pure and hybrid racial types. Agassiz conjectured that intergroup mixture produced racial degeneration, identifying human "hybrids" as inherently weaker or biologically unfit. Several other Age of Discovery explorers referenced similar terminology, including Tomlinson's flagrant use of the dehumanizing word "half-breed" (as cited in Wood and Chasteen, 2009). With studies like Darvasi's "Beauty of Admixture", we see specific applications of population mixture versus relative isolation, such as amidst the characterization of "admixed populations" as mean component percentages. The admixture concept creates a timeline of antecedent forms and subsequent intersection; it's a conceptual miscalculation, one that forces the researcher or

interpreter to think of genomic segments as having an original form within the "admixed" individual. The idea of admixture, and how it defines personhood, will be discussed in Chapter 3.

The invention and reinventions of scientific racism were never empirically founded. They were semantic and abstract, drawing upon decontextualized excerpts and misinterpretations of prominent scientific theories. Excitement surrounding the Human Genome Project elevated it as "one of mankind's greatest odysseys. It is a quest that is leading to a new understanding of what it means to be a human being" (Hauskeller, 2013, p.875). Although both a painstaking and monumental milestone, the Human Genome Project raised numerous more questions than it answered—a healthy impact to have on the research community, even if perplexing. Nonetheless, discourse surrounding the HGP revitalized the ability of scientists, and specifically geneticists, to "speak authoritatively about what makes us who we are, challenging the position held by the social scientists" (Hauskeller, 2013, p.875). In her book, *The Social Life of DNA*, Alondra Nelson (2016) explores how society has shifted its perception of personal origin to lie not decisively with one's narrative voice or oral tradition but ultimately within one's "DNA as the final arbiter of truth and identity" (p.4). Relinquishing this role to DNA and its interpreters not only minimizes the social, interactional, and abstract truths we create but also lends to scientific misuse. Today's sociological thinkers, such as Barbara Katz Rothman and Troy Duster, have recognized continuity and cautioned against a resurgence in the form of genetic determinism, a phenomenon where "one's biological inheritance [is] believed to indelibly shape one's health, and other attributes" (Nelson, 2016, p.12). While genetic inheritance and gene-environment interactions shape our physiology, metabolism, and health, we only understand a small window of the vastly complex, molecular interactivity of our cells. There is danger, and hubris, in assuming predictive power over one's outcome based on a genetic profile. Genetic

determinism would provide a novel way to retroactively inform today's disparities and rationalize complicity by understanding genes to predispose life trajectories. But, genetics does not override context, personal growth and experience, individual autonomy, or the elusive step from corporeal to abstract and body to mind that we have yet to understand. Humans exist in an abstract world, in which we construct our own realities, milestones, credentials, transactions, interactions, identities, stories, and meaning. Nature is not on a tier higher than nurture. But, if we are not careful, giving authority to DNA as the "final arbiter of truth", we begin to lay the foundation for a resurgence of lessons still unlearned from previous centuries of colonialism, repression, and reliance on biological explanations for social processes.

CHAPTER FOUR

Genetic Admixture Mapping

A Close Look at Operationalized Personhood

Admixture mapping, a steadily advancing technique used to discern genomic ancestry, exemplifies the quandary of capturing personal origin in scientific rhetoric. Geneticists are walking a tenuous line between race and "genetic ancestry", "biogeographic ancestry", "continental ancestry", etc. They edge ever closer to outdated racial typologies with the conceptualization of admixture, and specifically of "admixed populations", which links groups of people to particular demographic histories; so far, the admixed populations they have derived include specific racial and ethnic communities whose heritage they describe in terms of continental ancestry. Prior to the development of admixture mapping, most genetic association studies primarily included participants of European descent with the idea that European ancestry, on average, produced genomes of longer linkage disequilibrium blocks that, in turn, required fewer resources and markers to analyze (Nievergelt et al., 2013). Geneticists then began to surmise that allelic differences within that margin of human genetic variation could potentially confound association studies, misleading geneticists to correlate disease prevalence with allele frequencies that are unrelated to the disease, a circumstance known as a false-positive association.

Geneticists today believe that the ability to control for population structure is a prerequisite for GWAS and is necessary to prevent false-positives (Nievergelt et al., 2013). Since false-positive error effects increase with sample size, Nievergelt et al. argue that large-scale GWAS, especially those studying multifactorial diseases, cannot ignore population structure

(Nievergelt et al., 2013). In response, geneticists supported that robust scientific research requires diversity of participants as well as population-specific research methods. Diversity is vital in any scientific study striving to represent humanity but is seldom prioritized. When geneticists set out to increase diversity, they meant genetic diversity rather than diversity of identity. Since there is no way to genetically screen donors for the unique polymorphisms they carry from past ancestral communities[16], the means of increasing diversity became tied to social identity. And, the social identities they tapped were those they predicted to have the most "divergent" ancestors. Translated into race and ethnicity, this meant they pivoted to African American, Latinx, and Indigenous communities. While the intention of increasing diversity makes sense at a conceptual level, the curation of who belongs to an "admixed population" is reminiscent of the early HGDP's targeting of Indigenous peoples and "untouched" lineages. Ironically, admixture mapping has become characterized by disproportionate rather than diverse focus. The theory of admixture mapping had to be actualized through real people—"admixed populations". This move, how it manifests in studies, and its potential implications will be discussed below.

Researchers view admixed populations as coincidentally useful because of their demographic histories, or the preceding life events that have presumably led to a detectable convergence of diverse lineages. In an applied context, admixture[17] can refer to a variety of events that, whether voluntarily or forcibly, brought together different communities. The historic enterprises as well as crimes against humanity that technically explain genetic admixture, as it

[16] The chronology dilemma is at play again. If we already knew which polymorphism or allele frequencies were upregulated in which ancestral communities, groups which may not even coincide with how we socially describe ancestry, admixture mapping would be an obsolete endeavor for geneticists.

[17] Recall that, conceptually, admixture refers to the intersection of different ancestries, described as parent populations. These parent populations often include broad continental groups (i.e African and European), and admixed populations descend from the "admixture" of two or more parent populations.

has been abstractly conceived, could include the following: the perpetration of slavery, colonialism and neocolonialism from the 15th through 19th centuries, state-building in the 20th century, and the effects of a more cosmopolitan modern era lending to a multi-ethnic and changing world (Garrod, 2006). Let's recall the foundational underpinning of admixture mapping from Chapter 1.

4.1 From Concept to Practice

The theory is that, for admixed populations, too few generations have passed since the beginning of admixture for allele frequencies to experience the random effects of genetic drift or restructuring events of migration, marriage, and descendant generations (Shriner, 2013). Theoretically, admixture confounds genetic studies, as expansive immigration and global admixture diminishes the primacy of geography as a determinant of genetic variation; however, it has been reappropriated as a strategic tool (Novembre & Peter, 2016). The admixed populations that geneticists are pursuing for studies are supposed to be descended from "deeply divergent ancestries" (p. R224), bearing haplotypes or other polymorphisms that are attributable to distinct parent populations (Henn et al., 2010). Ultimately, geneticists try to compare allelic patterns of ancestors through the genomes of their descendants (Henn et al., 2010). While the vast majority of gene variants exist everywhere, there are specific allele frequencies with notable differences. For example, within certain communities in Western Africa, the null Duffy antigen[18] is at a fixed frequency of 100% while it remains near 0% prevalent outside of this geographic

[18] The Duffy antigen binds to human red blood cells, while the absence of the Duffy antigen from red blood cells is described as the null Duffy antigen phenotype. The null Duffy antigen phenotype appears to relate to environmental pressures; in order to infect individuals, the malaria parasite *Plasmodium vivax* relies on the presence of the Duffy antigen (Howes et al., 2011). So, null Duffy antigen phenotypes generate resistance to infection. Geneticists have located the chromosomal location of the gene that creates the Duffy antigen as well as discerned four allelic variations and ten possible genotypes (Howes et al., 2011).

location (Shriner, 2013). When such stark proportions exist, if an individual carried the alleles for the null Duffy phenotype, a geneticist could reasonably infer that the individual likely had ancestors from one of the regions where the null phenotype is fixed. In fact, some studies have used the Duffy antigen to evaluate the degree of admixture in a sample of individuals to determine if they are suitable for MALD because of its known and distinctive allele frequencies (Smith & O'Brien, 2005).

Interestingly, Duffy antigen variants are often described as blood groups, which re-invokes racialized ideas of bloodedness. We should not confuse the null Duffy antigen as defining West-African communities or as a West-African gene. The causal genotype likely arose because it makes red blood cells resistant to the malaria parasite *Plasmodium vivax*, a deadly pathology that exists in several continents and countries (Howes et al, 2011; Howes et al., 2016). Genetic markers that are considered fully informative, which are 100% prevalent in one parent population and 0% in the other, exist in an ideal situation, but the human genome contains a minute number of fully informative markers (Shriner, 2013). Alleles become fixed or lost in unpredictable ways (i.e. random structuring effects of a bottleneck event or even a mutation that leads to disease immunity), and it is a very rare occurrence that can be simply due to chance; furthermore, the impacted allele frequencies can again be restructured relatively rapidly as individuals move into and out of the affected group. Therefore, it is challenging to know where to look for stark allelic proportions. More often, ancestral inferences in admixture mapping are tenuously probabilistic because almost all alleles are equivalently prevalent across parent populations (Shriner, 2013). Furthermore, trying to classify "deeply divergent ancestries" orients the convergence of peoples relative to a specific time period pre-globalization, marking an

antecedent timeframe as evolutionarily relevant while using modern nationalities to describe ancestral groups.

4.2 Itemizing and Decontextualizing Identity

So, through the lens of genomics, who represents admixture of divergent ancestries? African American and Latin designations are the two most prominently identified and studied "admixed populations" to date, alongside which geneticists are currently trying to decipher classification criteria for other potential "admixed populations", such as among Native Americans and the First Nations of Canada (Grind et al., 2019; Henn et al., 2010; Shriner, 2013; Smith & O'Brien, 2005; Verdu et al., 2014). For example, Nievergelt et al. (2013) summarize the progress of concurrent admixture and GWAS studies in constructing marker panels "for Hispanic populations, African Americans, or three-way admixture in the Americas" (p.2) alongside other panels dedicated to global samples. While geneticists project admixture to become increasingly ubiquitous over time, they understand certain "major population groups, such as Latino or Hispanic," to "already represent an admixture of ancestry" today (Bonham et al., 2018 p.1534). In essence, they view Latinx communities as notably admixed due to the long-term impacts of colonialism and slavery that brought together African, Indigenous, and European peoples in Latin America as well as subsequent nationalism and the attempted integration of distinct communities under a national identity. In their study on "Interethnic Admixture and the Evolution of Latin American Populations", Fransisco Mauro Salzano and Mónica Sans (2014) provide an extensive examination of the regional histories, national identities, and genetic data of Latin American communities, regions, and countries. Their study is among the most comprehensive and historically informative review articles included in this thesis, but the central focus remains classifying each population and/or region of Latin America according to a gross

breakdown of percent European, African, and "Amerindian" ancestry (Salzano & Sans, 2014). Even more commonly than Latin American communities, African Americans have been designated as the primary admixture population-of-interest (Shriner, 2013). Millions were robbed from their families, clans, and tribes, of their freedom and lives, as a result of the Middle Passage that forced African peoples into the Americas. Despite being the reason for admixture, these histories are largely erased from the methods sections of genetic studies.

In order to systematize admixture studies, geneticists have essentially itemized the genomic inheritance of admixed populations. In doing so, they implement a rhetoric of scientific neutrality that appears to eclipse social and cultural significance. Researchers define African Americans as admixed according to the following parameters: they project that admixture began approximately 8 generations ago, and about 50,000 random markers, ranging from 39,000-160,000 total, are needed to map chromosomal segments inherited from different ancestries (Shriner, 2013). For individuals with Latin American roots, Shriner's "Overview of Admixture" (2013) estimates admixture beginning about 16 generations prior to today, requiring significantly more genetic markers to map ancestral differentiation. Additionally, admixture mapping necessitates a genetic map, which denotes local recombination rates, or the probability recombination will occur and ancestry will switch at a given locus; a genetic map specific to African Americans was reported by Hinch et al. in 2011, and project databases such as 1000 Genomes and HapMap have been used to construct these maps (as cited in Shriner, 2013). Based on these generational parameters, geneticists have also tried to deduce and reduce complex heritages into percent breakdowns, or continental "admixture proportions" (Nievergelt et al., 2013, p.2). The admixture of people with Latin American roots is commonly described as 50% European and 50% Indigenous in genetic origin with variable percentages of African heritage,

specifically in the Caribbean and regions in South America (Smith & O'Brien, 2005). African American ancestral lineages are commonly termed "approximately 80% African and 20% European in genetic origin" (Smith & O'Brien, 2005, p.4). Henn et al. (2010) cite a slightly different gross breakdown of 75% African descent and 25% European descent as well as sometimes 5% Native American descent. Within these gross breakdowns, we not only see through a reductive lens of what constitutes personhood, but we also conflate the "elaborate system based on the family, the lineage, the clan" (Deng, 1997, para. 3), the distinct ethnicities, languages, values and institutions within Africa, the different Indigenous roots and confederations of groups, into one single, vast continent. Commonplace notions of admixed African Americans are not not Afrocentric but reinforce a colonial construction of Africa, leveling it as comparable in scale and scope, in terms of both ethnicity and potential genetic diversity, to Europe. A similar admonition can be applied to Latin America, impacted by a complex history of colonialism, displacement, and enslavement as well as modern race relations. Although geneticists are scrutinizing African American and Latinx communities, admixture can apply to any intersection of groups, such as between the ethnic groups of India (IGVdb Consortium, 2005). This calls into question how scale, or different levels of organization across different time spans, impacts the concept of admixture. How do geneticists discern whether their population distinctions actually represent groups with differing genetic substructure? How rapidly do transient differences again restructure after reintroduction, or "admixture", between groups that diverged in the past hundreds or thousands of years versus tens or hundreds of thousands? These questions rely upon far-reaching theories of human dispersal, which continue to be scientifically debated as new fossil, ancient DNA, and archaeological evidence emerges.

Individual variation inevitably defies the gross percentages applied to admixed populations. Geneticists do note that, regardless of reported averages, any African American individual could vary from 1-90% in terms of European ancestry (the same concept applies to African, European, and/or Indigenous origin of an individual with Latin American roots), elucidating in scientific terms that these admixed population parameters are generalizations of mean data and neither override individual variability nor the unique intricacy of one's own heritage (Smith & O'Brien, 2005). Yet, despite acknowledging these technical limitations, the operational definitions used for admixed populations normalize the oversimplification of personhood, history, and heritage as relative percentages. This discourse also disguises race in a scientific context, framing ancestry through raw percentages that proxy grave histories and processes of racialization that lent to perceptions of "recent admixture". Impassive language is common, and often beneficial, in scientific literature, a consequence of the fundamental aim to prevent confirmation bias and rely as much as possible upon discovery through repeated observation. It is both practical and important for scientists to create operational definitions, delimitations, and assumptions for their experimental variables to make the studies understandable and enable others in the field to replicate their methodology; however, in the numerous circumstances where natural and social sciences intersect, modelling variables can mask the real life implications, or explanations, surrounding the research. In the context of admixture mapping, evading historical depth about the evolutionary parameters believed to have impacted a group of people not only desensitizes the realities of African American and Latinx admixed heritage but also could paradoxically be a detriment to study design by overlooking variables or misconceiving scale.

When we decontextualize race, we engage in a collective amnesia that insulates science from history, collective memory, and social realities. When discussing heritage as admixed, as percentages, we are often invoking a grimmer history. Paul Brodwin (2002), a professor of anthropology specializing in medical humanities at the University of Wisconsin-Milwaukee, unveils this tension, often overlooked by genetic studies: "Many people have complex mixed genealogies, created by sexual exploitation and the deliberate mixing of enslaved Africans during the Middle Passage and on American or Caribbean plantations" (p.327). Dr. Rick Kittles has anecdotally explored the realities of genetic admixture by sharing his own personal genealogical findings in his lectures. Through mtDNA analysis, Kittles speculated potential relations to the Hausa of Nigeria, but he follows this revelation with the "sobering news that his paternal Y-DNA traced to Germany" (Nelson, 2016, p.34). Kittles goes onto attribute this result to what he describes as the "'Thomas Jefferson effect,' gesturing at once to the sexual violence of slavery and to the DNA analysis that, along with archival records, strongly suggests the third US president fathered a child with Sally Hemings, a woman he enslaved" (Nelson, 2016, p.34). Kittles captures the pervasiveness of sexual exploitation by white Americans as evident even among presidents and preeminent figures, champions of liberty, still revered today; caricaturing systemic violence as the "Thomas Jefferson effect" is especially salient considering his maxim of "all men" being "created equal", thereby juxtaposing purported American values against the perpetuation of human bondage. A systematic review of racism and sexual health of African American women from 1619-2018 cited that approximately "58% of all enslaved women aged 15-30 years were sexually assaulted by slave owners and other white men" and were barred from legal protection against white men due to their statutory designation as property (Prather et al., 2018). And, white slave owners often held sickeningly deliberate intentions of fathering more

children into slavery without losing capital (Prather et al., 2018). These social realities underlie characterization of genetic admixture as relative percentages, molding criminal acts into scientific terms that allow researchers to bypass the conversation. We must also be wary that common names used for groups, as well as reimaginations of heritage, are often reproductions, adaptations or preservations of the "ways in which European Americans historically have defined those others than of how they have defined themselves" (Foster & Sharp, 2002, p.848). The population labels used by genetic admixture studies, including but not limited to "African-American", "Latino/a", "Hispanic", and "Native American", dilute the innumerable possible branching ancestries that can precede different people who socially or culturally identify similarly, overcasting the reality of vast individual variability by the systemizing shadow of percentage metrics. Casual uses of percentages in scientific contexts not only "[conceals] a great deal of cultural, linguistic, and biological variation" (Foster & Sharp, 2002, p.847) but also reflect terms colloquially used among the public to describe race or ethnicity and could unwittingly invoke a perceived link between biology and race. Uncritical disseminations of study results, whether to the general public or within applied settings (i.e. public health), could reify imaginations of racial typology and associated social stigma.

4.3 The Risk of Stigmatization

An individual's DNA, by nature, implicates other people, whether those genetically related, the communities they are from, or the identities they carry. Critics of genetic ancestry caution sampling from "smaller" populations, especially marginalized communities that have "experienced the disadvantages of minority status within larger polities", often in spite of composing global majorities (Foster & Sharp, 2002, p.847). The prioritization of the majority,

whether explicitly or implicitly, renders a double-edged unifying and "othering" effect that neglects needs of communities that are not routinely considered within the normative societal narrative. Humans, especially groups of humans, are more likely to maintain a conceptual numbness to experiences that are neither first-person nor prominently represented in media, policy, education, or scientific research unless they assume an intellectual responsibility of researching and engaging with diverse perspectives. Any community diverging from the default norm, but especially those smaller in number, experience disproportionate vulnerability to social stigmatization (Foster & Sharp, 2002). Foster and Sharp (2002) respond to these patterns by asserting:

> ...smaller social populations (particularly those that already are economically or politically disadvantaged) should not be identified in genomic resources or publications unless there is the potential for direct benefits to those populations, such as identifying genetic variants that predispose members to disease…. (p.847)

How does one weigh possible downstream benefits of research participation for their commmunity against the potential to expose more than just the donors, but also entire social groups, to stigma or discrimination depending on how the data and sample categories are extrapolated? Regardless of population size, numerous social identities experience disenfranchisement, whether socially, politically, or economically, due to processes of racialization and entrenched systemic barriers from histories of formerly and formally legitimized discrimination. Concerns of genetic donors are often "informed by historical experiences of mistreatment and exploitation by outside researchers" (Foster & Sharp, 2002, p.847). As discussed in Chapter 2, the United States has integrated suppositions of biological identity into policy, deeming Native American or African American legal identities via

rudimentary ideas of phenotype and ancestry (Foster, 2002, p.848). Indigenous and colonized communities are among the most susceptible to scientific hubris, which, in the realm of modern genomics, could look like the invocation of ancestral haplotype maps to predict or "confirm" ethnicity in lawsuits today (Foster & Sharp, 2002). As an analogy for the potential implications of admixture mapping and ancestral inference in the United States, we can look to a national forensics database in the United Kingdom.

The National DNA Database in Britain (NDNAD) is among the largest forensics archives in the world. The database is created, and updated, from biological material recovered at crime scenes and contains DNA information on ~4.9 million individuals by late 2009, including those over 10 years of age arrested in relation to any recordable offense as well as those who have been acquitted (Wallace, 2011). Freedom of Information requests by GeneWatch UK in 2006 revealed that controversial and nonconsensual research unrelated to open criminal cases had been conducted using the samples from the NDNAD (Wallace, 2011). This research included the use of Y-chromosome DNA to develop methods of predicting ethnic appearance based on genomic information, tapping into "both the DNA profiles on the computer Database and the stored DNA samples" (Wallace, 2011, p.89). These backdoor, auxiliary research endeavors are especially concerning given the disproportionate number of young Black men, and Black individuals in general, who are stored in the bank. Biobanks, and geneticists who use them, maintain a role in curating the demographic image of a society, and the NDNAD is undoubtedly refracted. Within the United Kingdom, about 47% of the Black population, 42% of all Black men, and 77% of all young Black men have had their DNA recorded in the NDNAD compared to a meager 6% of the entire white population; these figures were approximated by comparing those identified as "Afro-Caribbean" by police with the number of those self-identifying as an equivalent ethnic

category on the prior census (Wallace, 2011). Such skewed representations of a citizenry are bound to more frequently yield skewed results, whether in the field of forensics or adjunct genetic research studies.

Living amidst pervasive institutional trends of racial bias, potential volunteers might find themselves contending with how their participation in a genetic study, or contribution to a database project, might become intertwined with sociopolitical applications of genetics research or frame social difference within medical fields (Foster & Sharp, 2002). But medicine is not the only area where DNA research, and the sensitive information it contains, can resurface. Hauskeller et al. (2013) shed light on the social, cultural, and political potency of DNA, often looked to as a source of (seemingly) objective truth:

> Many social institutions, including not just science and medicine but also insurance companies, employers, government departments, policing and childcare services, all seek to ascribe status and identity using DNA tests… test data are in effect made to matter as determinants of identity. (p.878)

This potentiality raises three crucial points of contention; first, we must be attuned to how genetic research, the unpacking of current findings and technical limitations, is being positioned as an objective truth. The amount of scientific and social weight that we endow genes informs the second point, which is how institutions are linking genes to identity or taking one step further by conceiving identity as determinative of individual health, or even social, trajectories. Finally, we must question how the implication of admixture mapping in applied research fields, like public health and epidemiology, both implicitly and explicitly preserve concepts of biological race. Blell and Hunter (2019) caution the stigmatizing potential of operationalizing human populations in clinical settings:

> Reinscription of the notion of biological race in medical consultation, even inadvertently, validates the idea that race and ethnicity are natural classifications and runs the risk of encouraging racial/ethnic stereotypes and oversimplifications of the complex origins of most disease, leading to both a naïve genetic essentialism and a misunderstanding of human genetic diversity in society (Lee et al., 2008; Race, Ethnicity, and Genetics Working Group, 2005). In addition, the effective disregarding of social production of health models of understanding the ethnic patterning of ill-health, which have been developed on the basis of epidemiological research over the past decades, in favor of an assumption of a genetic basis, leaves ethnic groups' own supposedly shared faulty genes as the supposed cause of their ill health. (p.4)

This argument hones in on the essentialist logic that follows the prioritization of ancestry as a determinant of health risk, which diverts focus from social systems and scapegoats intergroup genetics. MALD specifically sieves for genetic markers of specific ancestries, placing certain groups under a microscope by nature of methodological convenience. While genes continue to gain momentum in applied research, a concurring phenomenon has led to the appropriation of genes, and semblances of genetic identity, as sociopolitical tools. The crossroads between genes, ancestry, and identity is already being reclaimed by political activists, root-seekers, and reconciliation projects striving to leverage genes to uplift communities (Brodwin, 2002). The social and political potency of DNA will be further discussed in Chapter 6, which examines the intersection of the genome with identity, policy, and concepts of personal origin. While such projects can be powerful, they contend with the complex relationship between identity, linked to how we personally experience and navigate the world, and genes, a raw code lacking inherent social meaning.

Most geneticists invoke social identities because they believe them to be the most accessible predictors, not conclusive indicators, of common genetic substructure; but, this fact is not explicit in most primary literature. Editorials and review articles are more likely to indicate this key property. Geneticists are nonetheless reporting statistically powerful analyses using complex computations and growing genetic marker banks, opening a dangerous door of misinterpretation. Ideally, medical professionals would understand social identities as a stand-in that, at most, hints toward one's probability of inheriting or developing a disease; however, implicating sociocultural identities as informative of disease etiology may lend to gross misuse via both public interpretation and private enterprise, news outlets, insurance companies, the unregulated supplement industry, and even pharmaceutical companies. June 23, 2005 marked the FDA approval of the New Drug Application (NDA) for BiDil, a modern form of racialized medicine (Kahn, 2011). NitroMed, a Massachusetts biotech company, markets BiDil as a heart failure drug specifically for African American patients due to its prior efficacy in a clinical sample group of self-identified African Americans (Kahn, 2011). Race-based claims have since been challenged due to the fact that no other demographic group was included in the African American Heart Failure Trial (A-HeFT) and that the pill was a combination of hydralazine and isosorbide dinitrate (H/I), two preexisting non-racial vasodilators for heart failure (Kahn, 2011). And yet, BiDil, which did appear to reduce mortality by about 43% in its clinical trial, attained federal approval (Kahn, 2011), gaining a legal and intellectual authority that could prompt, or be leveraged as evidence for, interpretations of race as a biological category. A concurring fact that BiDil sheds light on is that the vast majority of clinical trials for drugs marketed to everyone used to predominantly test white men, insinuating that "'white' was coextensive with the category 'human being'" (Kahn, 2011, p.130). So, when a drug was tested in only African

American subjects, it somehow was concluded as solely fit for African Americans. Kahn (2011) explains the subliminal message of racialized medicine:

> This sends the unintended but nonetheless powerful message that [B]lack people are somehow less fully representative of humanity than are white people. (p.130)

Furthermore, Kahn (2011) poses the essential question of who constitutes an African American for the purpose of this medication. Would efficacy diminish depending on the degree of different racial identities in one's ancestry? These questions could have been relevant if BiDil actually proved more efficacious African Americans, and even though BiDil is a modern form of legally supported pseudo-science, the inevitable emergence of these questions reifies race as a biologically important as well as reinforces a reductive concept of ancestry as proportional.

There can be strong impetus, whether scientifically, economically, or otherwise, to jump from new ideas to research applications, outpacing progress before building a solid foundation of basic knowledge or assured methodology. This tendency especially relates to studies of genetic ancestry and the potential for racialized applications. Foster and Sharp (2002) recommend that genetic researchers take proactive measures to speak up about limitations of their studies by routinely "[framing] subsequent public uses of race and ethnicity in relation to genetic features" (p.848) prior to publishing literature to the public eye. Alongside severing notions of biological race, we should also be careful of the creation of new identifiers and constructs, analogous to HIV positive communities, and how the routinization of genetic testing might create unprecedented markers of stigmatization (Foster & Sharp, 2002). Proactive measures and methodological reevaluations may help reduce the rapid spread of misinformation about genetics, which truly is promising as well as fascinating science (and has many different focuses aside from human population genetics). The fact that DNA is an intricate natural code, or

scientific language, inherited by all living things leads us to contemplate the secrets it can reveal about the history of life, about evolution, and about identity. But, as we have seen, the bridge between biology and identity is anything but simple. If genetic research more actively incorporated historical context, and discussions surrounding, both researchers and audiences might be more sensitive to how they handle population classifications, discuss racial and ethnic groups (i.e. African American), evaluate the sources of genetic samples (i.e. the NDNAD), define ethical criteria, or critically examine data.

The general public has acclimated to perceiving genes as the holy grail of scientific discovery, but this esteem lends to a complementary potential for grave misuse. Human beings are composites of multiple identities, some inherited and others formed but all constructed to some degree, and this reigns true for our branching lineages; concepts of identity will only continue to transform in a global neighborhood. Yet, we see in modern genomics, and specifically admixture, a fixation on empirical difference. While Francisco Mauro Salzano and Mónica Sans (2014) provide a better genetic research model by relaying historical context and explaining its scientific relevance in their interethnic admixture study, they do characterize Latin American populations as "natural experiments" for examining the genetic basis of "unique anthropological and epidemiological issues" (p.151). Criollo-Rayo et al. (2018) similarly offer that populations in the Colombian Andes "[represent] an opportunity to study admixture dynamics" (p.e1) of Indigenous and European heritage. There is a continuity throughout time, and throughout scientific history, of exposing and dispossessing Indigenous peoples, of studying the perceived "mixing" of whiteness with Black and Brown communities. Although the science is vastly different, there is continuity in the scientific imagination of race, and this idea again resurfaces with admixture mapping. Concepts like admixture mapping steer us to think in terms

of racial purity versus mixing. Admixture positions European and African continents as whole and African Americans as composed of parts—European and African.

From 19th century travelogues to admixture mapping today, scientific subdisciplines continue to designate entire communities as specimens of disproportionate scrutiny. How can scientists adapt a more critical lens and sensitive language as well as dismantle reifications of racial typology and gross misinterpretations? As Michael Yudell (2011) asserts that:

> ...despite the best intentions… to reconceptualize the concept of race for modern biology, evidence suggests that these geneticists and their scientific allies ultimately helped to preserve the concept of race in science, and hence for use by both scientific and nonscientific racists. (p.21)

This trend is noteworthy in prominent spheres, such as European countries moving to extricate race and insert discourses on "ethnic tensions" or UNESCO cautioned race as a "social myth" (as cited in Yudell, 2011, p.22) in its 1950-1 "Statements on Race" but also opted for ethnic designations as a replacement (Azarmandi, 2017). The racist uses Yudell (2011) references include those by William Shockley, a Nobel Prize recipient, professor, and physicist, who asked the National Academies of Science to ascertain a genetic explanation for the "slum problem" he perceived in America, wholly ignoring the structural inequalities that pervade his society and invoking Social Darwinist causality. In reality, geneticists should only include participants if they can affirm their best interest in the research intent. Perhaps data should not be released nor popularized that is preliminary and still contends the inherent flaws of categorizing social identities for the purpose of scientific study. Amidst the limitations of genetic ancestry, concepts of heredity, genealogy, DNA, and genetics have an evolving "social life", as termed by Alondra Nelson (2016), and significant role in cultural concepts of personal origin. As different people

increasingly look to genetic science for validation in social and political spheres for different reasons, we should look critically into the validity of the tool itself, not just technical and statistical nuances but translational limitations that lie at the intersection of genes and identity. Chapter 5 will further discuss the quandaries of genetic ancestry in academic research, specifically within the biomedical field and disease-susceptibility research, focusing on different limitations that arise during the progression from concept to methodology to practical application.

CHAPTER FIVE

Genetic Ancestry is Socially Constructed

Translational Limitations of GWAS and MALD

In spite of the overwhelming genetic similarity present across all human genomes, we continue to fixate on the science of difference. While the tools and techniques of genomics have significantly advanced since the Human Genome Project, a critical evaluation of the genetically-derived "human population" remains both crucial and neglected. Population genetics must be understood as well as explained as a probabilistic science, one that infers conclusions from correlations between and within predefined groups. Within all fields of science, but especially when working with probability, defining limitations is essential for preventing gross misinterpretation. After researching several academic studies from the past two decades, I conceptualized two broad categories of limitations; the first set involves analytical barriers, or those attached to the mapping techniques, statistical metrics, and computational models or programs geneticists choose when conceiving their methodologies; various analytical limitations were briefly overviewed in Chapter 1, but each iteration of study design presents its own nuanced capacity for error. The second set of limitations are translational, or emerge amidst the jump from concept to practice, statistic to inference, or sample group to human patients. The discourse surrounding admixture mapping, discussed in Chapter 4, has covered one facet of the most notable translational limitation, the construction of the "human population." Geneticists are struggling to streamline classification schematics and delineate human populations, but they continue to report associations with ancestral groups that are, for the most part, socioculturally

defined. Another significant barrier, especially when looking to create medicines, therapies, or preventative approaches in the biomedical sphere, is the potential to underestimate environmental factors in prioritizing genetic substructure. While both sets of limitations retain important contingencies, the conceptual leaps that implicate sociocultural identities involve tenuous logic that can potentially put entire communities at risk, opposing the principles of reliability and consistency that are central to the scientific method.

A review of how current studies define ancestry, or genetic ancestry, will help clarify the translational limitations of classifying human populations. First, most researchers conceptualize genetic ancestry at the level of the genome rather than the individual. Due to the repeating process of recombination that occurs with each generation, the genome fragments into an amalgam of chromosomal regions, or the termed "mosaic of segments" (Royal et al., 2010, p.665) that each bear independent inheritances. In some articles, such as those by Rosenberg et al. (2002) or Batai and Kittles (2013), the macroscopic concept of continental ancestry is readily used to describe people today as descendants of broad continental populations; the resolution is high-level, and this version of ancestral organization is informed by theories of ancient human dispersal to and from around five main continental regions (Royal, 2011). Chapter 3 discussed how continental groupings often fit outmoded models of human racial typology, which broadly classify humans as fundamentally African, American, European, Asian, and Australian using typological terminology (i.e. Caucasoid). Biogeographical ancestry (BGA) is also regionally-bound but might subdivide major continents; BGA typically considers other demographic qualifiers, including ethnicity, in addition to locale. Moreover, the study of fine-scale population

structure[19], or intra-country classifications, opened the door to conflations of ethnicity, religious affiliation, nationality, among other sociocultural identities.

While several concurring terms exist, neither the terms nor their definitions seem to remain consistent, diminishing the reliability of results, comparisons across studies, and extrapolations of data into the real world. In their article on the "red herring" of genetic ancestry, Blell and Hunter (2019) cite different instances in which primary articles have grappled with the concept of ancestry. They denote that, within the study "Genetic Structure of Human Populations", Rosenberg et al. reference equivocal variations of ancestry:

> One clear example of this comes from the landmark Rosenberg et al. paper on the structure of human populations (Rosenberg et al., 2002). In their paper, the authors utilize a number of different conceptions of the term "ancestry" without providing any sort of robust operationalizable definition of the term. Included in the article are "self-reported ancestry," "genetic ancestry," "Mongol ancestry," "self-reported population ancestry," and "genetically inferred ancestry" (Rosenberg et al., 2002). (Blell & Hunter, 2019, p.3)

Importantly, the paper lacks transparency, failing to define how each term is used in the study as well as its inherent assumptions. Blell and Hunter (2019) also relay how some researchers have differentiated levels and subdivisions within the umbrella of genetic ancestry, such as "*geographical* (i.e., south-east Asian vs. northern European); *geopolitical* (i.e., Cambodian vs. Swedish); and *cultural* (i.e., Jewish vs. Berber)" (p.3). Since DNA databases do not capture information beyond one or two generations ago (although studies that have analyzed DNA

[19] Geneticists such as Novembre & Peter (2016) agree that studying fine-scale population structure is contingent on the designation of individual origin and the sampling methods of "location, individual birthplace, or an origin based on parental or grandparental ancestry" (Novembre & Peter, 2016, p.102). They attest that, due to the scope of human migration, selecting a definition is important for sound study design and can complicate the interpretation of results (Novembre & Peter, 2016).

remnants from burial grounds do exist), genetic studies derive human populations from people living today and infer origin through comparisons with other "contemporary populations" (Royal et al., 2011, p.661). For example, in admixture studies centering African Americans, such as one by Bhatia et al. (2011), HapMap participant donors from Yoruba (totalling 113 individuals in this study) are frequently proxied as the African parent population to serve as a point of genomic comparison (Royal et al. 2011). This facet complicates the data that geneticists have accumulated on ancestral populations and the pool of knowledge from which they continue to build. Ultimately, geneticists have not been able to evade the act of socially constructing the "human population", and the implications of this logical fallacy is apparent throughout modern analyses. The remainder of this chapter will center the translational limitations of today's academic research pool, the majority of which is purposed with studying human genetic disease, discerning ancestral disease risk, or accumulating basic knowledge to augment these goals.

5.1 Classification Conflations and Inconsistencies

Currently used human population classifications, which are instrumental to reference databases, admixture mapping, genome-wide association studies, or informativeness statistics, are structurally incongruent. While the study of human populations relies upon the existence of scientifically relevant ancestral groups, no streamlined method of distinguishing populations exists. Earlier studies adopted continental ancestry as an acceptable modality but primarily implemented it as a proxy for colonial understandings of race, such as the use of the term "Caucasian"[20] alongside "African" or "Asian" in a 2004 AIM patent publication (Frudakis, 2004). Beyond continental ancestry, the advancement of fine-scale population structure as well as increased mapping resolution have only caused classifiers to become more convoluted. In an

[20] Again, the use of the term Caucasian invokes Johann Blumbach's five human varieties as well as later adaptations of his theory.

earlier article on “Human Races and Evolutionary Medicine”, Swynghedauw (2003) problematizes equivocal conflactions of race in scientific contexts, as researchers will assign “several, and contradictory definitions” (p.438) that entangle geographical origins, imaginary constructs (i.e. Caucasian), and religious affiliations. Consistency of variables is central to the scientific method; a field of study should maintain the ability to operationally define and derive relevant participant samples that not only reasonably reflect the individual makeup and mean characteristics of the population that scientists are trying to vicariously study but also are defined and selected in the same manner across all studies that are comparing the same metric (i.e. genetic ancestry or population structure). The selection of imprecise metrics might curate genetically-arbitrary groups and misattribute relative genetic similarity and difference, and the relevancy of groupings likely fluctuate based on the allele being studied; furthermore, the geographical regions that researchers gerrymander and cross-compare may not contain equivalent genetic variance.

These problems exist because geneticists sample people rather than genomes or genotypes; in order to study interpopulation difference, they need a way to extrinsically assign people to different human populations. We can see a lack of consensus surrounding the “human population” across studies. For example, Henn et al. (2010) cite West Africa as among the most genetically diverse regions in the world. The International HapMap Consortium (2004) also distinguishes communities in Africa as having “more genetic variation than other world populations” (p.473). Yet, continental Africa is often diluted as a genetically-relevant population classification. Classification schematics also oscillate from study to study, complicating the ability of researchers to replicate study design and build up the reliability of conjectured interpopulation comparisons. Rosenberg et al. (2003) try to compute informativeness data from

52 different populations, each organized into intra-region designations, visible in table 1, while Henn et al. visualize the "population structure of worldwide human populations" as seen in figure 3:

Table 1

Comparison of Correlation Coefficients Across World and Population Data

Data Set	Description of Groups	Correlation Coefficient		
		I_n and *ORCA*	I_n and I_a	ORCA and I_a
World-52	52 populations representing seven regions	.920	...	...
World-5	5 regional groups (Africa, Eurasia, East Asia, Oceania, America)	.956	.994	.956
World-7	7 regional groups (Africa, Europe, Middle East, Central/South Asia, East Asia, Oceania, America)	.953	.990	.947
Africa	6 populations (Bantu [Kenya], Mandenka, Yoruba, San, Mbuti Pygmy, Biaka Pygmy)	.945	.986	.937
Europe	8 populations (Orcadian, Adygei, Russian, Basque, French, Italian, Sardinian, Tuscan)	.878	.973	.867
Middle East	4 populations (Mozabite, Bedouin, Druze, Palestinian)	.873	.994	.891
Central/South Asia	9 populations (Balochi, Brahui, Makrani, Sindhi, Pathan, Burusho, Hazara, Uygur, Kalash)	.883	...	...
East Asia	18 populations (Han, Han [N. China], Dai, Daur, Hezhen, Lahu, Miao, Oroqen, She, Tujia, Tu, Xibo, Yi, Mongola, Naxi, Cambodian, Japanese, Yakut)	.915	...	...
Oceania	2 populations (Melanesian, Papuan)	.921	.998	.940
America	5 populations (Karitiana, Surui, Colombian, Maya, Pima)	.934	.989	.945

Note. This table reports the informativeness of microsatellite markers in data subsets, separated into regional affiliations of the 52 constituent populations that were genotyped. The correlation coefficients of informativeness for assignment (I_n), optimal rate of correct assignment (ORCA), and informativeness for ancestry coefficient (I_a) are compared for each subset of data. Reprinted from "Informativeness of Genetic Markers for Inference of Ancestry," by Rosenberg et al., 2003, *The American Journal of Human Genetics*, *73*(6), p.1410.

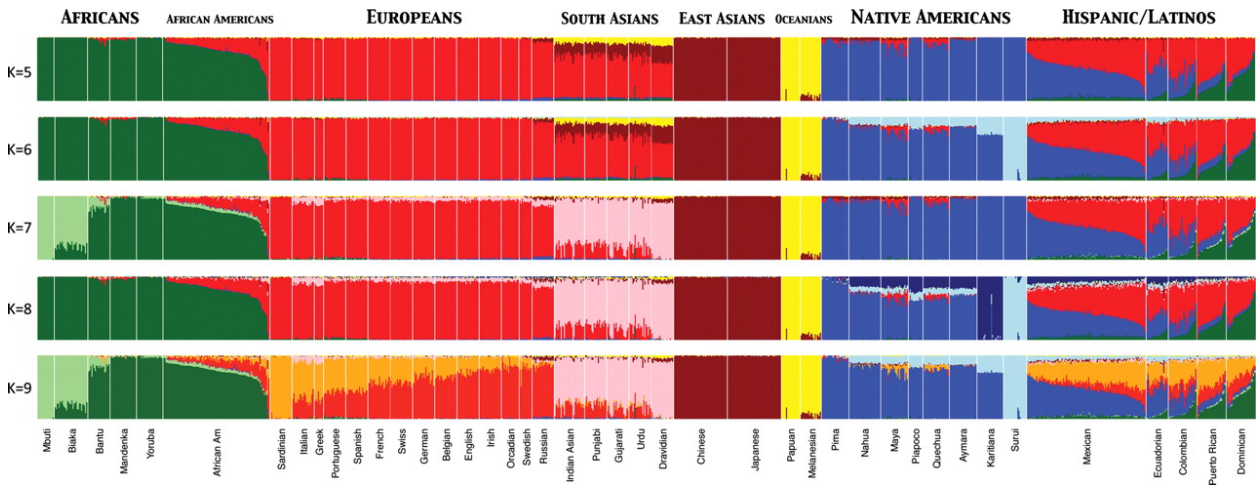


Figure 3. Visualization of global population structure from compiled individual ancestry proportions of 1112 participants and 42 populations. Individuals are represented by vertical lines on each barplot; it is apparent that Europeans comprised the majority of the participant pool. A clustering algorithm in the computer program ADMIXTURE was used to stratify global genome estimates from K=5 to K=9, which refers to the number of ancestral populations inferred for each individual. As the number of ancestral populations inferred increases, Henn et al. attest that intracontinental fine-scale structure becomes visible. Reprinted from "Fine-scale Population Structure and The Era of Next-generation Sequencing," by Henn et al., 2010, *Human Molecular Genetics, 19*(R2), p.R222.

Taking a closer look at the group descriptions and the figure axes, several properties stand out. A conceptual inconsistency is apparent on the upper x-axis of figure 3 between "Hispanic/Latinos", an ethnic designation, with "Europeans", "South Asians", and other geographic delimitations; "Hispanic/Latinos" appears to serve as a proxy for Latin America but includes a limited set of identities, each defined by country. Although circumscribed, the inclusion of five linguistic/ethnic groups under the "Africans" label juxtaposes both the vast

subdivision of “Europeans” and the generalization of African Americans as a single, comprehensive group, simplifying complex ancestries as scientifically commensurate. If geneticists are trying to implicate histories of slavery and colonialism as relevant events in structuring differential ancestry, then they are neglecting the full story, not only by omitting their reasoning but also by failing to name all the other regions to where enslaved peoples were forcibly moved (i.e. the perpetration of slavery by Spain in Latin America, most notably Brazil). Interestingly, the greatest number of intrapopulation designations fall under “Europeans” despite previously cited logic that West Africa alone likely maintains the most genetic diversity among geographic regions (Henn et al., 2010). Overall, the intra-region designations that are included in the group descriptions of table 1 and listed on the lower x-axis of figure 3 fluctuate between ethno-linguistic, religious, and cultural constructs as well as nationalities, and the researchers appear to grapple with a lens superimposed by the formation of nation-states.

These subcategories both simplify and conflate complex demographic histories, levelling each to a seemingly, but erroneously, comparable level of organization. For instance, apparent in both images are the classifications “Bantu” and “Italian”. While “Bantu” refers to the Bantu languages, which sustains a complex history of expansion throughout Eastern and Southeastern Africa since 1000 BC, as well as ethnic and indigenous communities today, “Italian” refers to diasporic identities from a nation-state defined in 1861 (Isern & Fort, 2019; Viola & Verheul, 2019). Viola & Verheul (2019) attest that the “Italian” national identity was a “vacant [abstraction]” (p.295) and had to be formed both socially and politically; it is tied to the Italian language, but dilutes its “politically, culturally, economically, and linguistically” (p.295) distinctive constituencies that were abruptly conflated amidst the making of a unified Italian state. The inconsistencies of the aforementioned studies exemplify one of the fundamental

problems of invoking social constructs of identity, which evolve across context, place, and time, under the guise of scientific validity.

5.2 A Refracted Image: Identity Assignment and Self-Reporting in Human Population Genetics

While the conflation of dissimilar identifiers presents inherent problems, the process of assigning research participants to human populations adds another layer of complexity, as geneticists primarily rely upon self-reporting. In his critical analysis of modern racialized science, Garrod (2006) describes today's world as increasingly cosmopolitan, lending to the intricacy and intentionality of self-identification. He relayed an important distinction made by Dr. John E. Clark, a Professor of Pharmacotherapeutics, former president of the Association of Black Health-System Pharmacists (ABHP), and advocate of culturally competent health care systems (FSHP, 2019): "For me to call someone 'black' because they [have] black skin may be inappropriate because they may not see themselves as that" (at cited in Garrod, 2006, p.57). The conceptual paradigm of race, or the version that has been normalized by past colonial discourses and present census checkboxes, is being reimagined, characterized by a growing rift between self-identification and outside perception; the infamously pervasive "other" category is ironic given the insufficiency of the alternative blanket categories. This dynamic, among other factors, pokes holes in the reliability of self-identification. Social constructs are relevant, real, and dynamic, but they cannot be systematized as a classification schematic. Nonetheless, they are the methodological fulcrum of ancestral inference and, as we will see, critically complicate human population genetics.

Here, the question of chronology becomes central, as most studies perform analysis using the DNA samples provided and labeled by reference databases. In their viewpoint editorial on

ancestry and biomedical research, Bonham et al. (2018) summarize the intended purpose of sociocultural identities as " surrogates for ancestral background" (p.1533) in genetic studies. There are two degrees of separation between the evolutionary processes that may have helped shape an individual's genome and self-reported race, ethnicity, or other identities. The logic is as follows: first, self-reported identities have to be relatively predictive of ancestral heritage, by which I mean the geographic origin and movement of ancestors, the communities to which ancestors belonged, or the locations different ancestors intersected (or encroached). Ideally, they reflect some semblance of the demographic history preceding an individual; however, the predictive ability of any given self-reported identity is associative, unable to reveal causality, and is not, in and of itself, indicative of ancestral heritage (Bonham et al., 2018). Second, the ancestral heritage preceding an individual, specifically in regard to geographical origin, must be a pertinent factor to the genetic substructure that exists today. The relative contribution of geographical ancestry to modern genomic differences has likely been diminished by vast intersections of peoples throughout history, since population differentiation relies upon the relative isolation of communities over evolutionarily significant timespans (Garrod, 2006).

Nonetheless, researchers are motivated by the potentiality that our distant human history has left a genetic imprint enabling them to derive clinically relevant populations, and they continue to leverage self-identification as their compass. In their studies on ancestral inference, Nievergelt et al. (2013) and Pardo-Seco et al. (2014) corroborate that self-reported ancestry is sufficiently reliable for predicting how a sample population will continentally cluster. Yet, contemporary studies consistently fail to report the specifics of their identification criteria, or whether and how participants were instructed to self-report. Were they asked to report race, ethnicity, or nationality; consider context or perceptions of others; or indicate the geographic

origin, nationality, race, or ethnicity of their grandparents? These criteria might be outlined in the databases they use, but criteria across databases varies. If we take a step back to the semantic character of the questions themselves, we can investigate an inconsistency that circumvents the reliability testing evidenced by contemporary geneticists.

The United States Census offers an interesting point of comparison for discussing data collection of self-identified race and ethnicity. Two decades ago, the US Census encountered methodological barriers, as approximately seven million respondents identified with more than one race (Swynghedauw, 2003). This reality underscores the importance of data collection metrics, especially if survey questions circumscribe rather than accommodate the ability to define multiple identities. In any qualitative assessment, the phrasing of a question, as well as the format of the answers, can influence how an individual chooses to respond (Strmic-Pawl, 2018). This basic problem of data being framed by questions and answers problematizes census derivations of race or ethnicity, as individuals hold multiple identities shaped by personal experience, family ties, cultural heritage, citizenships, among other factors. Strmic-Pawl et al. (2018) discuss how a respondent might contend with census self-reporting:

> Respondents may see these terms as overlapping and redundant, or they may see them as independent and thus have multiple identities. For example, someone who is Puerto Rican could identify as Puerto Rican as an ethnicity, race, nationality, and ancestry. Or, someone who is Puerto Rican could identify with a Hispanic ethnicity, a White race, an American nationality, and a Taino ancestry. Thus, the terms used in the phrasing of the question can determine the identity provided by a respondent. Moreover, relying on the word "race" in the question does not necessarily inform as to which identity is most salient or important to that person's lived identity. (p.6)

Foster and Sharp (2002) link this phenomenon to the prospect of identifying DNA donors in a Genome Research commentary article:

> An individual donor, for instance, may be known simultaneously as a resident of a particular Indian village in Arizona, a member of the Hopi tribe, a descendant of a Laguna family (through a paternal ancestor who is not explicitly noted in matrilineal Hopi society), a Native American, and as someone of Spanish ancestry (owing to 18th-century intermarriages between Lagunas and Spaniards).... (p. 847)

Race is imagined in the context of science but real in the form of experience, and the dynamic ways in which someone navigates the world, faces barriers or prejudices, and finds community are all components of race as well as ethnicity, nationality, ancestry and numerous other lived identities. The modern nation-state further complicates the question, as national identities can sometimes superimpose upon or be referred to as ethnicity or ancestry. This tendency is historically rooted in the evolving lens of identity. Various ideas of race have been constructed in different contexts, as we have seen in past uses of today's religious-cultural, ethnic, and national markers as distinct races; while the process of racialization evolves overtime and according to place, race concepts also produce durable elements and ideas that persist through multiple centuries. Importantly, while race can be imagined and reimagined in relation to different peoples, the systems of oppression that have racialized people and the consequences of these systems are not equivalent across communities, and certain communities can eventually become perceived as assimilated into the normative culture. Whiteness, which became meaningful during colonization, was oriented as the norm while non-white identities were racialized and othered, as literally inscribed by the "other" category on many surveys. Factors such as the dynamic nature

and living history of racial constructs influence how respondents conceptualize and answer questions regarding their own race, ethnicity, ancestry, or identity.

The challenges of US Census data collection help elucidate the limitations of modern population genetics. We must direct more focus on how DNA samples are sourced, or the inclusion criteria used by genetic database projects and sample collection methods of primary studies. Among the most prominent reference databases to date, the International HapMap Project defines a "population" as "a group of people with a shared ancestry and therefore shared history and pattern of geographical migration" (International HapMap Consortium, 2004, p.469). The International HapMap Consortium (2004) note in an aside that major scientific and ethical contentions exist within the field, rooted in the fact that individuals hold multiple identities or the circumstance where individuals who identify with a specific community do not reflect the ancestral history of the group average. In their first phase methodology, HapMap sought out specific populations to sample; certain choices were influenced by the desire of national funding agencies to include their own majority population and others based on practicality (i.e. researchers at Howard University and the University of Ibadan had already established a partnership at the time as well as created a trustful relationship with the Aba Alamu community in Ibadan in Nigeria) (International Hapmap Consortium, 2004). Subsequently, they created inclusion criteria to confirm membership in each population, which participant donors had to meet based on self-identification (International Hapmap Consortium, 2004). The International HapMap Consortium (2004) described the inconsistencies of their first phase inclusion criteria, recognizing the limitations of self-reporting ancestry of oneself as well as of parents or grandparents. They yielded that "as in many population genetics studies" the methods of demarcating and sampling from the "populations themselves… were inexact" (International

HapMap Consortium, 2004, box 2, p.469). In this research phase, the HapMap Project employed slightly different inclusion criteria depending on the population, asking Han Chinese participants to identify "at least three Han grandparents" and Yoruba donors to report "four Yoruba grandparents" but omitting a questions about familial origin when speaking with the participants from Japan in an attempt to maintain cultural awareness and avoid direct questions about parentage; instead, they prefaced to all potential participants before they chose to donate that the aim of the study was to collect genetic samples from individuals whose grandparents were from Japan (International HapMap Consortium, 2004, box 2, p. 469). Inconsistencies such as these render interpretation of data and generalization of population structure tenuous. It is difficult to conclusively affirm whether the data acquired from these particular participants are generalizable beyond the specific donor sample to a broader population, and, if so, how individuals would know or report that they meet a threshold level of genetic similarity with a given population. Large-scale projects, like HapMap, that are trying to ascertain global populations and archive every available faction of genetic diversity run into significant problems surrounding the equivalency of groupings, conflation of categorically different sociocultural identities, and inconsistency of self-reporting metrics; regional projects have followed suit, more precisely studying fine-scale population structure but face paralleling limitations.

The field of human population genetics pivoted to the study of "populations" on a finer scale within countries and regions despite its fundamental sampling problems remaining unsolved. Several regional database projects have been developed concurrently, each tackling the intricacy of intra-country cultural, religious, linguistic, and socioeconomic signifiers. In "The Indian Genome Variation database (IGVdb)", human populations were identified on the basis of both geographical boundaries and linguistic families by a group of anthropologists and social and

community health workers with the help of fluent-speakers of local the languages (Indian Genome Variation Consortium, 2005). The degree of endogamy in family genealogies was qualified through interviews and pedigrees (Indian Genome Variation Consortium, 2005). The researchers disseminated a questionnaire that included designations for ethnicity and family disease history as well as phenotypic traits; although the questionnaire acquired multiple identifiers and descriptions of marriage pedigrees, information was not provided detailing the language of the questions themselves, the approach of the translators, anthropologists, and social/community health workers, or the accommodation of persons who could not read or fill out the informed consent (Indian Genome Variation Consortium, 2005). While multidimensional recordings of identity are better than generalizations, the context and potential influencers of self-reported metrics remain impactful contingencies regardless of the number of identification metrics provided.

The "Iranian Human Genome Project" parallels features of the IGVdb, as well as other regional database projects, purposed with surveying the human genomes of Iranian ethnic identities, described as differing in culture, lifestyle, languages, and spatial distribution, for implications in national health care throughout the country (Banihashemi, 2009). Like other concurrent genome projects, this study outlines far-reaching goals of "illuminating our understanding of Iranian ethnicities' history and identity" as well as "creating a unique bridge between science and the humanities in Iran" (Banihashemi, 2009, p.89). Interestingly, the study reports that extensive social, historical, linguistic, and cultural knowledge was necessary to accumulate enough data for the studied populations to be initially identified, taking the researchers over two years and involving local leaders in the design of sampling protocol for different communities (Banihashemi, 2009). This study appears to achieve a level of cultural

competency and intentionality unmatched by macroscopic projects, building extensive contextual reports rather than relying on the randomness of convenience; however, they fail to relay how ethnic identities were confirmed, or reported, by the participant donors beyond listing that parental birthplace was recorded and a coding method was used to categorize sample DNA by geographical and demographic criteria. If this study is to be replicated, which would help demonstrate the reliability of results, subsequent researchers would have difficulty maintaining consistency without knowing these details. Both between and within these fine-scale database projects, we can take note of apparent incongruities in design and focus. The IGVdb (2005) placed specific emphasis on endogamous marriage practices while the Iranian Human Genome Project (2009) used "language as the major criterion" (p.89), followed by religious-cultural markers. Comparing and contrasting the methods between projects of different scales illustrates the variability of deriving human "populations" for empirical study and determining who fits these prescribed frameworks.

A threshold level of transparency in inclusion criteria is necessary to understand not only the intended purpose of a study and the individuals studied, but also to understand limitations, determine comparability across studies, and assess whether there is a reliable way to generalize the results. Often, databases that compile sample DNA are caught in the same binds as the census as when they rely upon self-reporting as a way to mark or confirm the complexity of identity. Further, census data collects identities such as race to trace disparities and allocate resources not because race is a biological entity but because it is real in its social and systemic consequences. Although contemporary studies try to frame population metrics in terms of ethnicity, linguistic, religious, or other anthropological markers intended to hint at similarities in ancestral heritage

and descendance from communities affected by similar evolutionary processes, concepts of race are caught within this matrix of data.

What does DNA tell us? Race? Nationality? "Ancestry"? Actually, none of the above.

It is not that DNA indicates, or can conclusively profile, any of these constructs vis-à-vis nucleotide sequences. The chronology is the converse in that geneticists use constructs created and sustained by human communities to hint at, predict, or proxy a higher likelihood that an individual's ancestors had something in common in terms of geographic movement or relative isolation and duration with others identifying similarly. Even if such qualifiers turn out to be substantive predictors of concurrent histories (and even so, predictive ability would decrease over time given our increasingly global society), they could never be a 1:1 ratio, as sociocultural constructs are not causal of underlying genetic variance in any individual. Thus, the census analogy and closer look into project methodologies raise central questions about how geneticists retrieve information on an individual's ancestry. Interestingly, after the 1960s, the Census Bureau shifted from identity assignment to self-identification because 1) one's own concept of self is "the most accurate since anyone else's perception is a guess (dependent on perceiver's experience and social context)" (p.6) and 2) race is a social construct and "not biological" so "there is no correct answer that can be assigned" (Strmic-Pawl, 2018, p.6). The genetics community likely standardized self-identification for similar reasons, but the information they are retrieving is only a guide for an underlying genetic population metric that remains elusive. They are not using sociocultural identities as social scientists might, noting their limitations and implications as social constructs, but deriving associations between genes and peoples as well as

conjecturing differential disease susceptibility based on these identities. Again, Strmic-Pawl et al. (2018) provide illuminating insight:

> The racial and ethnic data with which we work are never a true or accurate reflection of society but rather a reflection of society as refracted through the categories provided. (p.7)

They are, in short, imagined or constructed by individuals themselves in the midst of a vast array of identities. Geneticists must understand their role in "refracting" the genetic population map they are working to create. In the realm of forensics, we have observed the role of refraction manifest in the coding system of the National DNA Database (NDNAD) in Britain (Wallace, 2011). The construction of these categories insinuates which communities the UK perceives as most important in the context of criminality. These categories are externally assigned to individuals and include Afro-Caribbean, Arab, Asian, Dark-skinned European, Oriental, and White-skinned European (Wallace, 2011), groupings of uneven resolution and outmoded language (i.e. "Oriental") that are disproportionately specific in profiling Arabic and Afro-Caribbean individuals. In addition to curating the image we see of genetic ancestry, geneticists must also be aware of the potential reductive effect of self-reporting surveys in perpetuating stereotypes of "mixed races", as if individuals possess fractions of each of their multi-racial or ethnic identities or that "pure" races exist and can be mixed. We cannot simply sieve multidimensional identities into reductive categories of incongruent resolution and assume the validity and reliability of their cross-comparison. Further, geneticists in current studies are failing to consider and capture all facets of identity, reporting beyond self-reported race and ethnicity to also include socioeconomic status and other environmental and social conditions; these facets should be included in the association analyses and concurrently visualized. Extrinsic covariates

are significantly important in epidemiology and biomedical research, and the overreliance on discerning ancestral groups for clinical purposes can result in underestimations of non-genetic factors.

5.3 Genes, Environment, and Human Disease

One of the primary reasons modern academic research focuses on studying genetic ancestry is to inform disease etiology. Even studies that solely evaluate ancestral informativeness statistics, discern population clusters, or map by admixture disequilibrium are often doing so to serve downstream benefits of understanding differential disease risk by ancestry or controlling for population structure. It follows that human disease research is one of the primary realms impacted by the translational limitations and potential misattributions of genetic ancestry. The interplay of genetic and non-genetic factors is pivotal to disease-susceptibility, and studies involving public health should never ignore the environmental, social, and economic factors that contribute to disease risk. Most directly confounding to genetic analysis is the relationship between the environment and gene expression. Nature presents a conflicting natural phenomenon, which Adams (2008) concisely summarize:

> A striking example of the power of gene regulation is seen in agouti mice, in which genetically identical twins can look entirely different in both color and size. For example, one mouse may be small and brown, but her twin sister may be obese and yellow. Another genetically identical sister may have a mottled look with both fur colors present and fall in the middle of the weight range. (para. 1)

Why is it that the same genome can branch into such starkly divergent outcomes? While our DNA encodes all of the instructions needed for growth and development, our genes are not static blueprints. Their embedded information can be edited over time through the regulation of DNA

itself or the molecules it creates. In fact, certain aspects of our physical appearance and physiological attributes stem from these nuanced changes, impacted by a branching array of extrinsic circumstances that impact us from childhood through adulthood.

During the process of gene expression, the cell's molecular tools interpret the instructions encoded by different genes and build proteins or other products, based on these instructions, that are ultimately responsible for expressing phenotypes. Several steps intercede this process, including both molecular and environmental mediators, and our bodies are in a constant and complex relationship with the outside world. Newer genetics research is focusing on gene-environment interactions, which describe how external stimuli can cause genotypes to express different variations of traits, which seldom have only one possible version (Portela & Esteller, 2010). In concert, the burgeoning field of epigenetics dissects the molecular mechanisms that can structurally and functionally alter an individual's DNA (Byrd & Hughey, 2015; Portela & Esteller, 2010). This genetic malleability is possible because of a network of molecules that coexists with cellular DNA and has the ability to activate or suppress genes entirely (Adams, 2008). Some epigenetic changes in the form or function of an individual's gene can be heritable, or passed down in its altered state (Byrd & Hughey, 2015; Portela & Esteller, 2010). Basically, gene-environment interactions describe the relationship between environmental factors and genetic loci, catalyzing changes that are often materialized by epigenetic mechanisms.

Unsurprisingly, the paradoxical uniqueness of "identical" twins exemplifies these mechanisms. Human monozygotic twins possess indistinguishable DNA sequences but show different molecular patterns of epigenetic modification, some of which can impact the development of diseases later in life (Portela & Esteller, 2010). And environment does not simply mean climate or geography but involves numerous external influences, such as social

experiences or psychological stresses; the relationship between epigenetics, social adversity, chronic stress, and brain plasticity is currently being researched to chip away at the age-old question of nature-versus-nurture (Notterman, 2015; Papadopoulos, 2011). Gene-environment interactions and epigenetics contribute to a more dynamic understanding of genetics than ever before, especially in the context of disease-susceptibility research (Byrd & Hughey, 2015). These processes are among the most prominent confounders of admixture mapping and other genetic ancestry techniques (Shriner, 2013). Importantly, gene-environment research is a relatively new field; researchers have not yet reached consensus on the environment's sphere of influence (Boardman et al., 2013). Moreover, the field faces many intricate dilemmas in terms of how to best operationalize environmental factors, distinguish individual characteristics from social contexts, and "account for group-level behavioral, normative, and cultural processes that shape individual health and behavior" (Boardman et al., 2013, p.S65). Environmental health determinants are not a simple equation of cause and effect but a widely variable array of relationships up- and downstream the multivariable path from environment to outcome. Ultimately, researchers are discovering that genetics are not as predeterministic as previously thought. Even on a genetic level, we can understand that our development is, to some degree, inexorably intertwined with our surroundings and circumstance.

The past two decades have witnessed improvements in statistically powerful computational models, genome-wide admixture mapping, database size, and other genotyping tools. Especially relevant to the clinical application of genetic ancestry, the study of epigenetic triggers has linked the environment to the development of cancerous, neurological, and autoimmune diseases as well as other complex pathologies, such as cardiovascular disease (Portela, 2010), as displayed in the following table:

Table 2

Epigenetic Mechanisms And Modifications in Different Human Disease Classes

Aberrant epigenetic mark	Alteration	Consequences	Examples of genes affected and/or resulting disease
Cancer			
DNA methylation	CpG island hypermethylation	Transcription repression	*MLH1* (colon, endometrium, stomach[11]), *BRCA1* (breast, ovary[11]), *MGMT* (several tumor types[11]), *p16^{INK4a}* (colon[11])
	CpG island hypomethylation	Transcription activation	*MASPIN* (pancreas[92]), *S100P* (pancreas[92]), *SNCG* (breast and ovary[92]), *MAGE* (melanomas[92])
	CpG island shore hypermethylation	Transcription repression	*HOXA2* (colon[20]),*GATA2* (colon[20])
	Repetitive sequences hypomethylation	Transposition, recombination genomic instability	*L1* (ref. 11), *IAP*[11], *Sat2* (ref. 107)
Histone modification	Loss of H3 and H4 acetylation	Transcription repression	*p21^{WAF1}* (also known as *CDKN1A*)[11]
	Loss of H3K4me3	Transcription repression	*HOX* genes
	Loss of H4K20me3	Loss of heterochromatic structure	*Sat2*, *D4Z4* (ref. 107)
	Gain of H3K9me and H3K27me3	Transcription repression	*CDKN2A*, *RASSF1* (refs. 115–116)
Nucleosome positioning	Silencing and/or mutation of remodeler subunits	Diverse, leading to oncogenic transformation	*BRG1*, *CHD5* (refs. 127–131)
	Aberrant recruitment of remodelers	Transcription repression	*PLM-RARa*[103] recruits NuRD
	Histone variants replacement	Diverse (promotion cell cycle/destabilization of chromosomal boundaries)	H2A.Z overexpression/loss
Neurological disorders			
DNA methylation	CpG island hypermethylation	Transcription repression	Alzheimer's disease (*NEP*)[135]
	CpG island hypomethylation	Transcription activation	Multiple sclerosis (*PADI2*)[135]
	Repetitive sequences aberrant methylation	Transposition, recombination genomic instability	ATRX syndrome (subtelomeric repeats)[135,143]
Histone modification	Aberrant acetylation	Diverse	Parkinson's and Huntington's diseases[135]
	Aberrant methylation	Diverse	Huntington's disease and Friedreich's ataxia[135]
	Aberrant phosphorylation	Diverse	Alzheimer's disease[135]
Nucleosome positioning	Misposition in trinucleotide repeats	Creation of a 'closed' chromatin domain	Congenital myotonic dystrophy[151]
Autoimmune diseases			
DNA methylation	CpG island hypermethylation	Transcription repression	Rheumatoid arthritis (*DR3*)[154,155]
	CpG island hypomethylation	Transcription activation	SLE (*PRF1*, *CD70*, *CD154*, *AIM2*)[6]
	Repetitive sequences aberrant methylation	Transposition, recombination genomic instability	ICF (*Sat2*, *Sat3*), rheumatoid arthritis (*L1*)[152,155]
Histone modification	Aberrant acetylation	Diverse	SLE (*CD154*, *IL10*, IFN-γ)[6]
	Aberrant methylation	Diverse	Diabetes type 1 (*CLTA4*, *IL6*)[159]
	Aberrant phosphorylation	Diverse	SLE (NF-κB targets)
Nucleosome positioning	SNPs in the 17q12-q21 region	Allele-specific differences in nucleosome distribution	Diabetes type 1 (*CLTA4*, *IL6*)
	Histone variants replacement	Interferes with proper remodeling	Rheumatoid arthritis (histone variant macroH2A at NF-κB targets)[157]

Note. This table outlines links between epigenetic triggers and genetic changes corresponding to different disease types, such as cancers, neurological disorders, and autoimmune diseases. Reprinted from "Epigenetic Modifications and Human disease," by Portela, A. and Esteller, M., 2010, *Nature Biotechnology*, *28*(10), p.1064.

Knowing that the interaction of environmental factors and genes can influence the onset of a disease during one's lifetime, we can understand how epigenetics challenges basic notions of inheritance and why geneticists caution extrapolating from single studies (Byrd & Hughey, 2015; Simmons, 2008).[21] Since these changes occur spontaneously, they do not directly reflect prior evolutionary processes or selective pressures that curate genetic probability. The potential for an individual's genetic expression to become altered and possibly be passed down to their children complicates searches for disease-causing genes, as GWAS relies upon detecting observable phenotypes of disease-incidence. The occurrence of these changes, independent of one's demographic history, also obscures associations between differential disease risk and ancestry.

Consider the following hypothetical instances of false-positive results. First, a particular sample of individuals living under parallel social, economic, or political circumstances might be similarly impacted by gene-environment interactions but are used in an association analysis as representative of an entire sociocultural identity or broader geographical population. Second, a particular demographic group could be disproportionately impacted by certain systemic factors and display a higher prevalence of a particular disease that is more likely connected to structural barriers than ancestry. These are two hypothetical situations, and many nuances exist when contextualized in individual experiences. When applied to admixture mapping, geneticists acknowledge that differences in disease-incidence between parent populations may arise from environmental factors, genetic risk factors, or their coupled effect (Smith & O'Brien, 2005). In their review of " Mapping by Admixture Linkage Disequilibrium: Advances, Limitations and Guidelines", Smith & O'Brien (2005) explain that MALD fails in a case where the frequencies

[21] Most common diseases are multifactorial, arising from the combined effect of genes, gene-gene interactions, environmental factors, gene-environment interactions, and non-genetic factors. Thus, phenotypic distribution does not inherently reveal underlying genetic structure.

of alleles associated with a multifactorial disease are equivalent between parent populations but disease-incidence differs between them due to social and environmental factors. In this circumstance, geneticists would not detect the alleles that are actually related to the disease-of-interest and might even misattribute the disease to unrelated alleles that appear to vary in frequency between study samples. Further, Portela & Esteller (2010) denote that epigenetic changes can enable a variety of possible phenotypic outcomes to stem from a single genotype. Remember, any given phenotype or trait arises from a combination of genetic makeup, environmental stimuli, and gene-environment interactions that impact an individual. (Shriner, 2013). As we already know, people who share similar phenotypes might not carry equivalent genotypes, but we also can identify that individuals with parallel genotypes might not actually possess the same phenotypic outcomes.

These complex relationships of gene, trait, and environment problematize studies that grossly correlate health disparities with racial and ethnic identities. A specific example of how public health research might misattribute causality can be gleaned from an article entitled "Impact of Race/Ethnicity and Social Determinants of Health on Diabetes Outcomes". Walker et. al. (2016) investigated relationships between race, ethnicity, and social determinants of diabetes to help the advancement of culturally-tailored medical care. Similarly to other public health literature, they used diluted categorizations—Asian Americans, Hispanic Americans, non-Hispanic black Americans, American Indians/Alaskan Natives, and non-Hispanic whites—that decontextualize the specific lived environments of different communities within these broad groups. While they tried to address both demographic groups and social determinants, details within their messaging insinuated that race and/or ethnicity has a directly causal impact on health outcomes, such as the indication that "minority populations… exhibit poorer self-management

abilities, and experience more diabetes-related complications compared to non-Hispanic Whites" (Walker et al., 2016, p.3). Kahn (2011) vies for us to take a critical lens, one that weighs the relativity of factors:

> ...what are we to make of the fact that African Americans suffer from disproportionately high rates of hypertension, but Africans in Nigeria have among the world's lowest rates of hypertension, far lower than the overwhelmingly white population of Germany? Genetics certainly plays a role in hypertension. But any role it plays in explaining such differences must surely be vanishingly small. (p.132)

Kahn's perspective alludes to the systemic role of structural inequality in perceived racial, or ancestry-based, health disparities. Walker et al. (2016) coalesce and describe important metadata on social determinants of diabetic health trajectories, covering a host of social factors, but the scientific narrative contains gaps in its derivations of group-based science. Health disparities like those described in Walker et al.'s study have been adapted as focuses of genetic research. In their study, "Race, Genetic Ancestry, and Health", Batai & Kittles (2013) shed more light on the correlations reported by Walker et al., attesting that "for hypertension, genetic ancestry does not appear to contribute much when you control for socioeconomic status (SES), and, as of yet, we have been unsuccessful in using admixture mapping to explain the higher rates of diabetes among African Americans" (p.6). An article by Cooper et al. (1997) provides further insight, explaining that while populations of African descent in North America, the Caribbean, and West Africa might appear to have genetic similarities, their BMI levels and the overall prevalence of hypertension vary extensively across these groups. This variability complicates simplistic prescriptions of causality as it relates to health disparities and genetic ancestry; Bonham et al. (2018) more generally caution that "race and ethnicity are operationalized inappropriately when

they serve as proxies for other demographic variables, such as an individual's socioeconomic status" (p.1533). If researchers detect associations between genetic or health-related variables and ancestry that actually correlate with differences in socioeconomic status, or other circumstances, misattribution of causality is probable.

While it is important to identify systematic disparities faced by marginalized communities, which can be visualized differently depending on what kind of information the researchers intend to derive, identity cannot be misinterpreted as the cause. The multidimensional nature of genetic disease-susceptibility research requires intentionality in the selection of variables and covariates most pertinent to a study's intended purpose. Yet, geneticists in current studies are failing to capture certain facets of identity with which disease risk could be simultaneously visualized, such as socioeconomic status, income bracket, residence location, or education level (Bonham et al., 2018). The first step toward tackling covariates is formally recognizing them, acquiring more data points from the samples-of-interest, such as mean data on the living conditions or sociopolitical climate of a particular study group. Furthermore, when considering the role of identity, of race, ethnicity, heritage, and personal origin, the question is multidimensional. Boham et al. (2018) delve into the scaffolding of self-concept, explaining that:

> ...other dimensions of race should be recognized, including perceived race or ethnicity (what others believe a person to be), reflected race (the race a person believes others assume [them] to be), and the cumulative burden of discrimination. (p.1533)

Studies that fail to account for the impact of race in our lived environment lack an important dimension of health and personhood. Although still an emerging realm of study, epigeneticists such as Aroke et al. (2019) are investing whether racial and ethnic health disparities relate to

psychosocial environments, formed by experiences of "childhood stress, racial discrimination, economic hardship, and depression" (p. 701) that can arise from structural injustices. Aroke et al. (2019) attest that DNA methylation, a specific process of epigenetic modification, occuring at a glucocorticoid (relates to hormonal stress responses) receptor gene called NR3C1 "has been associated with depression, childhood stress, low socioeconomic status, and chronic pain" (p.701). Prolonged states of chronic stress and pain have also appeared to correlate with methylation patterns in genes supporting the immune system (Aroke, 2019).[22] When race appears to surface as scientifically relevant, it is because of the associated challenges surrounding identity and the impact of these challenges on health. In his lecture series, Dr. Rick Kittles demonstrated an interplay between "histories of social isolation and discrimination" (p.33) and DNA, and even more specifically with racial health disparities and oncogenes[23] that regulate tumor cells (Nelson, 2016). Kittles' work on the African Ancestry DTC test, its mission, and its use in the reconciliation projects surrounding African American cultural identity and political discourse will be discussed further in Chapter 6. Racism is an oppressive system, and historic processes are not static but carry forth to modern day. Institutionalized barriers, such as the redlining of districts and refusal of housing loans to Black families to restrict them from suburbanization, continue to impact the social, economic, and political circumstances of African American communities today (Zenou & Boccard, 2000). In order to prevent generalized conjectures about race and help detect false-positive correlations with ancestry, multivariable

[22] Studies on the relationship between health, chronic stress, racism, and discrimination exists within public health research. Examples include Sawyer et al.'s (2012) study of physiological responses to anticipated discrimination or Harrell et al.'s (2003) evaluation of the relationship between racism and negative health outcomes. As seen with Aroke et al. (2019), epigenetics is also looking into how structural injustices (and states of chronic pain) relating to racism potentially influence epigenetic triggers. Both are intricate study focuses beyond the scope of this thesis that will likely continue to accumulate new approaches but also contend the complexities of causality and must be cautious of reifications of biological determinism or racial essentialism.

[23] Oncogenes are a specific class genes that can generate tumor cells when activated.

analysis that controls for the numerous factors that interplay is essential. Further, we now understand that extrinsic factors subject genetic associations to the confounding influence of epigenetics. And, when the primary community of focus is one that shoulders a history of persecution and continues to face resultant systemic barriers, we must especially scrutinize the role of environment, of social processes, and of structural inequality to not magnify race or ethnicity as more telling than context and continuity.

Health disparities can emerge from disproportionate care, economic access, or barriers in existing healthcare systems. Life experiences from childhood through adulthood, such as long-term nutritional status, treatment by providers, general workload, and the safety of one's living conditions, also impact health in interrelated ways (Fine et al., 2005). Multifactorial diseases are complicated by the interaction of genes, environment, and epigenetics. Therefore, truly informative studies must be constructed in a manner that, by capturing a multifaceted profile of individual donors, prevents the erasure risk factors (Smith & O'Brien, 2005). Additionally, genetics, among other scientific fields, are falling into an individualistic framework that focuses on curating medicine to individual profiles rather than sourcing and remedying causal factors (Byrd, 2015), such as mitigating environmental carcinogens[24]. While the ability to transform an individual's personal information into a tailored medical approach can be highly beneficial, we should not forget to also prioritize proactivity, allocating research and resources to the complex social, political, and economic scaffolding of health-related root causes. Such discussions are beyond the scope of this thesis but further corroborate the need for increasingly interdisciplinary research initiatives. For instance, in their study on gene-environment research methods,

[24] For example, there is an ensuing debate surrounding the airborne emissions of concentrated animal feeding operations (CAFOs) and its impact on the health trajectory of nearby residents (Von Essen & Auvermann, 2005). The pursuit of more research and regulations surrounding CAFOs is an example of tracking a root environmental health risk.

Boardman et al. (2013) argue for the benefit of framing future research with a social epidemiological perspective, providing new insights into the current pool of primary literature. Genetics could similarly benefit from reframing approach or inviting outside perspectives. Otherwise, the study of genetic ancestry, especially given the inconsistencies of population ascertainment, risks overlooking equivalent or greater, and potentially more mendable, spheres of public health research.

The translational limitations of this chapter problematize not only the conceptual basis of the "human population" but the practical application of genetic studies today. The inconsistencies of classification schema alone exemplify how race is not an inherent biological characteristic. In an applied biomedical setting, substituting race, ethnicity, and ancestry in place of critical health determinants (i.e. socioeconomic status), due to correlations between race and these variables, calls into question cause and effect - a fulcrum of the public health sphere. Proxying race, whether via demographics or geography, "opens the door to inequities in medical care" (Swynghedauw, 2003, p.439) and can impede other fields of study, including genetic research itself, that could catalyze the advancement of new and existing therapies. Furthermore, the concept of genetic ancestry, and the way it is operationalized as sociocultural identities, overshadows the institutionalized racism and cultural incompetency that contributes disproportionate access across groups under the pseudonym of race as a determinant (Swynghedauw, 2003). Through analysis of disease-susceptibility research, we can understand that real danger lies within the translation of results to reality, clinical settings, or public knowledge. Critics of genetic ancestry research caution the stigmatization of communities on the basis of genetic associations with circumscribed numbers of participants from tenuously defined ancestral categories (Fine et al., 2005). Genetic ancestry yields inherent inconsistencies and is

not a scientifically discrete variable, relying heavily on societal constructs of race and ethnicity today. Acknowledging race and ethnicity as socially-defined should not serve to delegitimize these identities but should reframe the reliance on heuristic, genetic data as a mechanism to prove ancestry, and thus identity as it relates to race or ethnicity through continental or other metrics of categorization (Batai, 2013). This notion of "proving" ancestry, or "discovering" roots, is not confined to academic research but has been popularized in public discourse surrounding personal origin. Chapter 6 will delve into the intricacies of ancestry and identity outside of academia as well as explore how commercially-available genetics has shaped conversations surrounding culturally-specific concepts of personal origin.

CHAPTER SIX

An Era of Commercialized Genetics

The Role of Genes in Concepts of Personal Origin

As individuals, each of us descends from an incredibly long, branching line of ancestors, most of whom remain unknown to us throughout our lifetimes. While we might "inherit" who our family is (whether biological or adopted), what we look like, where we are born, and how we grow up, our identity, as we come to know and understand it throughout life, is neither inherent nor static. Instead, identity is continually formed and reformed through the building blocks of circumstance, experience, and society. And, while we begin with unchosen circumstances, choice becomes fundamental to identity formation as we diverge from our starting point. This chapter will investigate the meaning of genealogy, the evolving social and political life of genetic ancestry, and contemporary sociocultural concepts of personal origin in the context of commercialized genomics.

6.1 The Genealogy: Ideas on Heritage and the Reciprocity of Individual and Community

Questions surrounding family, affiliation, and self often lead us to knowns and unknowns of the past. Our heritage, specifically in the form of genealogy, takes a fundamental role in answering the central question—"where do I come from?". The term genealogy itself also takes on a variety of definitions, from describing accounts of ascent to or descent from an individual to personal microhistories to the study of one's relationships and kinship (Hatton, 2019). Most of these understandings center how genealogy is used to explore concepts of personal origin, but genealogy also wields a social power as a tool for group affiliation. Several different concepts of genealogy, blood (both symbolic and scientific), and lineage have informed social, political, and

legal frameworks of group membership, and they continue to do so in evolving ways. But, the concept of the genealogy, and intersecting ideas of identity, is not a natural phenomenon but emerges from the human imagination. Genealogy is a way of organizing and visualizing kinship, thereby refracted by those who leverage it and according to their interests or goals; it is a social process, and the role of genealogy in identity formation assumes different social and political meaning depending upon place, time, and person. For example, some scholars argue that the "visual representation of a pedigree chart/family tree was important in the historical process of fixing birth as the beginning of kinship" (Hatton, 2019, p.6) rather than positioning kinship as crafted, earned, or endowed later in life, which can influence how genealogies are implicated in social contexts.

The individual and social purposes of the genealogy come together if we conceptualize identity as a reciprocity of self in relation to others; in other words, personal origin informs social belonging while group membership reinforces self-concept. In linking self to group, Hatton (2019) positions genealogy within separate contexts of "social, economic, occupational, and class membership" (p.8) and familial relationship. In "Community, Identity, and Cultural Space", Dismas Masolo merges these contexts into a single system. Masolo (2002) expresses reciprocity in terms of individual and community, referencing personal experience while denoting significant variability across cultures:

> Valuation of knowledge, including knowledge of personal and shared histories, sometimes differs significantly between cultures, and also between individuals within and across cultural boundaries. When and where I grew up among the Luo of Kenya, genealogical knowledge was important both in itself and for social and moral reasons. Knowledge of the larger social System of which one was part, and of one's exact location

> within it, was crucial for determining one's own and others' rights and duties as well as general customary comportment toward others. Individual and community were related in a constant mutual dependency: the specific behavior of individuals in various contexts gave the community its cultural boundaries and identity just as much as the normative standards of the community regulated the practices of individuals and groups within it. As one grows up and attains the age of adulthood in this cultural environment, this knowledge and the derivable behavioral expectations become more demanding. An adult Luo man or woman is always expected to behaviorally relate to others – by speech and deeds – within the limitations provided for (or expected of) the kinship relations between them. One knew or could know her or his relatives and calculate or adjust their behavior toward them accordingly. (p.22)

Masolo (2002) goes on to explain that a deep sense of kinship sustained the "Luo model" of social and moral order, or organization. Developing a sense of belonging often also produces a sense of obligation to one's community while also connecting the community to one's individual identity. Thinking about cultural heritage abstractly, we can also conceive culture as inherited only after it is formed by growing up in and/or practicing the culture of one's ancestors in the present, sustaining a collective memory that links members of the culture, or community, together by performing the past in the present. Hatton (2019) again provides a different angle on the role of kinship and lineage, explaining that the genealogical "tree preserved the memory of ancestors, enhanced the prestige of lineage, and sustained power (Klapisch-Zuber 1991, pp. 106–7)" (as cited in p.6). Genealogy, and specifically the social idea that kinship is inherited, could effectively ensure power and status was a function of family line. Again, genealogy is a way of

organizing kinship, and its role in personal origin and group affiliation is variable depending upon context.

Genealogy, or more broadly heritage, has been used to create group membership across antecedently arbitrary bounds, superimposing unity upon communities with vastly different cultures, values, languages, beliefs, and institutions. This process is especially prominent among the variety of state-building techniques implemented among nascent countries, which loosely extrapolated and conflated ideas of common ancestry with national identity. When Porfirio Díaz, the President of Mexico from 1877-1910, launched an initiative to excavate the Pyramid of the Sun at Teotihuacán, an Indigenous site commonly attributed to the Aztecs, he marketed the plan as an exploration and celebration of Mexico's Indigenous roots (Bueno, 2010). A proponent of neocolonialism and reverent of France as the model culture, Díaz hoped to attract outside investors as well bring order, progress, and a unified national identity post the 19th century Civil Wars (Bueno, 2010). This pursuit is interesting both because of its inherent irony as well as how Díaz and Leopoldo Batres, his commissioned archaeologist, leveraged ancestral root-finding (Bueno, 2010). See, to prepare the supposed celebration of Indigenous roots, Batres paradoxically expropriated land from and displaced the Yaqui and Maya living at Teotihuacán in the present day, arguing that he should not have to compensate considering that the site "belonged to the nation since the time of the Spaniards" (as cited in Bueno, 2010, p.63). Additionally, the state positioned the prestige of the Aztecs as the "Greeks of the yellow American race" (as cited in Bueno, 2010, p.71), attempting to achieve Eurocentric standards of ancient prestige and modern relevance. Díaz qualified the Aztecs in relation to the Greeks to position Mexico's ancestry in the same way Europeans claimed ancient Greeks as their cultural predecessor. Through Teotihuacán, Díaz attempted to craft visual proof of a national genealogy

that tied Mexico to “prestigious” and sophisticated Indigenous roots. This act of calling upon elusive, distant ancestries, creating ties to people who may or may not have, at some genealogical convergence, descended from the Aztec civilization, for decisive identity formation mirrors the shifting tone of ancestry today.

Genealogy, heritage, and identity are becoming increasingly tied to technology; genetic testing companies promote the purpose of their projects as “discovering” roots while the public augments their credibility by adopting the same discourse surrounding personal identity. More and more, people are reaching back through hundreds or thousands of years to explore what Hatton (2019) describes as “deep ancestry” (p.8), using apparent genetic affiliations to reframe present identities whether or not they actually share cultural traditions, familial bonds, or social realities of the corresponding groups today. Moreover, others are generalizing scientific theories of human origin to evidence sameness and override social disparities. Former US President Bill Clinton has been cited as expressing “unless your ancestors, every one of you, are 100 percent, 100 percent from sub-Saharan Africa, we are all mixed-race people” (Short, 2016). Here, Clinton refers to the out-of-Africa model, a theory of human dispersal commonly investigated by geneticists and facing continued debate, to assert a scientifically-backed colorblindness. Like Díaz, Clinton attempts to create a universal identity across every conceivable social boundary of human difference by drawing upon a shared notion of “deep ancestry”and, as it logically follows, multi-racial identity. Except, multi-racial identities exist because racialization created such distinctions in the first place. Clinton makes the mistake of overriding social reality with scientific “fact”, believing the theory simply and seamlessly negates race; by trying to level human racial identity, he erases the experiences and power systems that still exist, regardless of the science.

In contrast to Clinton's colorblindness and Díaz's unification strategy, there are instances in which genetic ancestries and genealogies have reaffirmed cultural tradition. While this use again elevates science as the "final arbiter of truth", we can see such instances to also be profound for the communities impacted. Brodwin et al. (2002) describes one such instance:

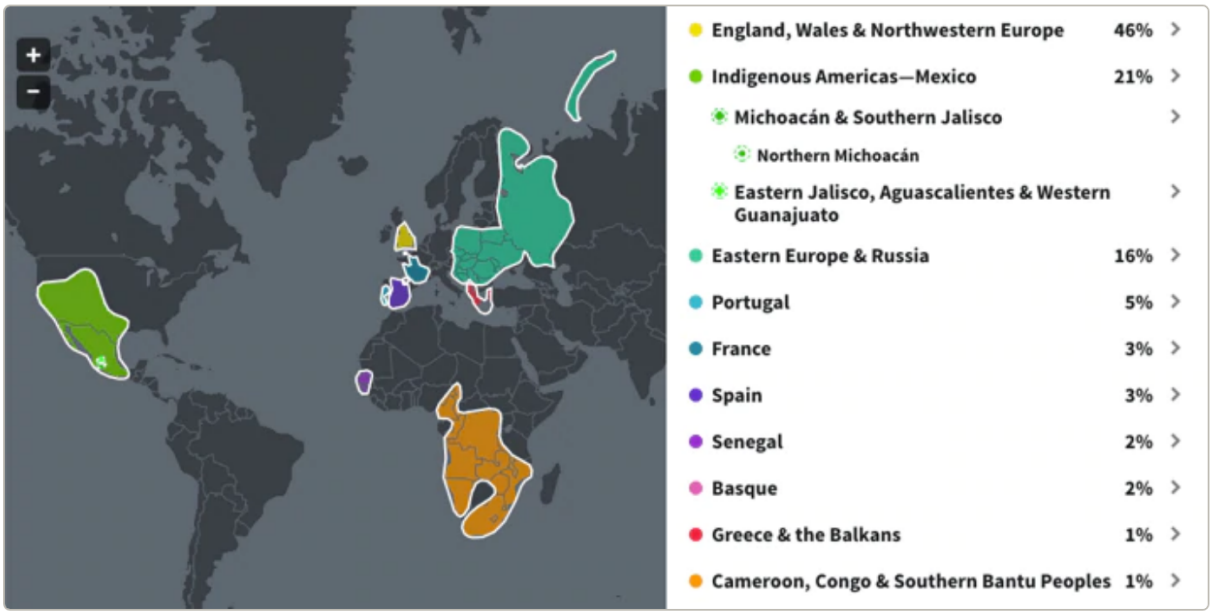
> ...geneticists in England have used Y-chromosome markers to demonstrate that at least one of the clans of Lemba, a tribe in South Africa and Zimbabwe, may be descended from Semitic peoples. Lemba interpreted the genetic findings as confirming their oral tradition of Jewish descent… [and] such practices as keeping one day of the week holy, circumcising newboard males, and not eating pork. However, what does it mean to say that this evidence 'confirms' the Jewish identity of Lemba? (p.325)

They prod at the unbalanced role of scientific evidence "confirming" identity, contemplating the political and legal implications of the statement. Brodwin et al. (2002) then pose the question: what if they had asked for citizenship to Israel, which offers a right of return to all Jewish peoples? In a separate instance, Isaiah Washington, an African American actor who took the African Ancestry DNA test during one of its early reveals, discerned genetic genealogical roots in Sierra Leone and was able to gain dual citizenship based on the results (Nelson, 2016). Interestingly, during the Ebola outbreak of 2014, Washington received criticism for carrying the Sierra Leone passport but not sharing the burden of the health emergency; in response, he lobbied at the UN for support and used the avenue of his philanthropy, the Gondobay Manga Foundation, to help provide relief and resources to Sierra Leone (Nelson, 2016). Brodwin et al. (2002) supposes that "to claim a certain social identity always implies certain rights and obligations" (p.325). But, how these rights and obligations will manifest amidst the proliferation of genetic ancestry is uncertain. The aforementioned two case studies of genetic ancestry alone

have drawn upon interrelating ideas of race, culture, religion, tradition, obligation, and citizenship. Such is the scope and manifesting social power of genetic ancestry.

6.2 A Technical Examination of Direct-To-Consumer Tests

The use of genetic ancestry as a tool to explore personal origin has become increasingly commonplace with the advent of direct-to-consumer (DTC) tests. Prior to commercial availability, genetic ancestry testing was typically confined to academic research. Now, the average consumer can purchase an at-home ancestry kit, swab their DNA, and mail their sample to a company laboratory, the central purpose being to trace ancestral origin.[25] In return, DTC companies provide consumers with a digitized report of their genetic genealogy, often describing these profiles as a " breakdown of your ethnicity by percentage" (*What to Expect from AncestryDNA®*, n.d., para. 3), as stated by Ancestry.com.



[25] DTC test companies also market separate genetic health kits to screen for disease-predisposing alleles, coming with their own sets of limitations and nuances, but this facet of commercial genomics will not be covered within this thesis.

Figure 4. Example of an Ancestry.com ethnicity report that shows the consumer percent compositions of personal origin subdivided by contemporary geographic states and regions. Reprinted from *What to Expect from AncestryDNA®*. (n.d.). Ancestry.com. Retrieved April 16, 2020, from https://support.ancestry.com/s/article/What-to-Expect-from-AncestryDNA

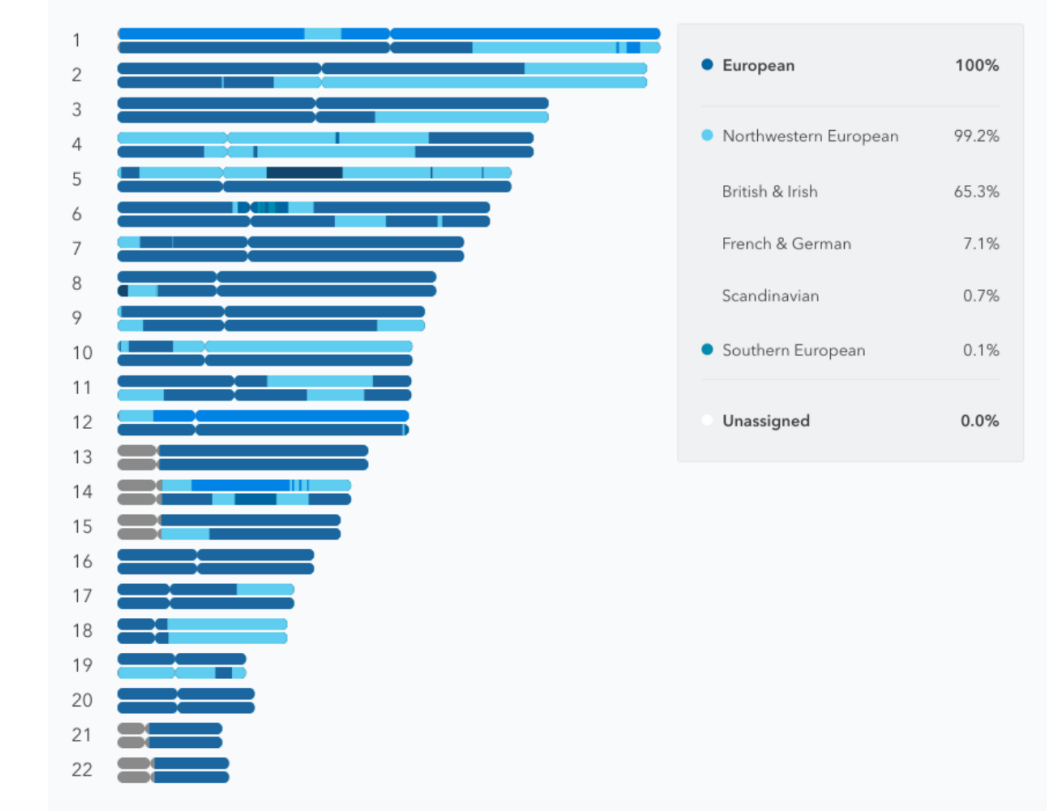


Figure 5. Sample 23andMe ancestry report visualized as chromosomal regions, which leverages the MALD concept that adjacent chromosomal segments possess different ancestral origins. This visualization reinforces the idea that chromosomal segments can actually be assigned ethnic values (i.e. European). Reprinted from *Ancestry composition*. (n.d.). 23andMe. Retrieved April 27, 2020, from https://medical.23andme.com/wp-content/uploads/2015/10/Ancestry Composition.pdf

Test results might visualize genetic origin using a map, emphasizing geography as a proxy for ethnicity, or display the genetic composition of one's chromosomes, building upon the conceptual foundation that human chromosomes comprise unique ancestral origins at different loci. In academic research, genetic ancestry typically serves a secondary function to the main research goal, unless the goal of a study is to assess or improve methods of ancestral inference for downstream benefits. Conversely, DTC ancestry kits serve to provide genetic genealogies, informing individuals of their scientifically-certified origin story.

Amidst the competing voices of the humanities, social science, and science, the general public routinely turns to the natural sciences for objective truth; in turn, there is potential for the validity and social jurisdiction of genetic genealogy to be overstated. The tenuous science and analytical limitations of DTC testing further complicates its role in identity, as the accuracy of DTC ancestry kits remains contentious among academic geneticists who caution that their limitations can skew results and mislead consumers (Pardo-Seco et al., 2014; Royal et al., 2011; Via et al., 2009). A drastically marked increase in the use of DTC tests has occurred over the past three years, ushering the consumer into a new era of data accessibility and commercial genotyping technology as well as adding new dimensions to the social process of identity formation. An article in the MIT Technology Review cited that about 29 million individuals had taken DTC tests by the start of 2019, noting an exponential upward trend forming around 2017 and the majority of tests taken with two companies—Ancestry.com and 23andMe (Regalado, 2019).

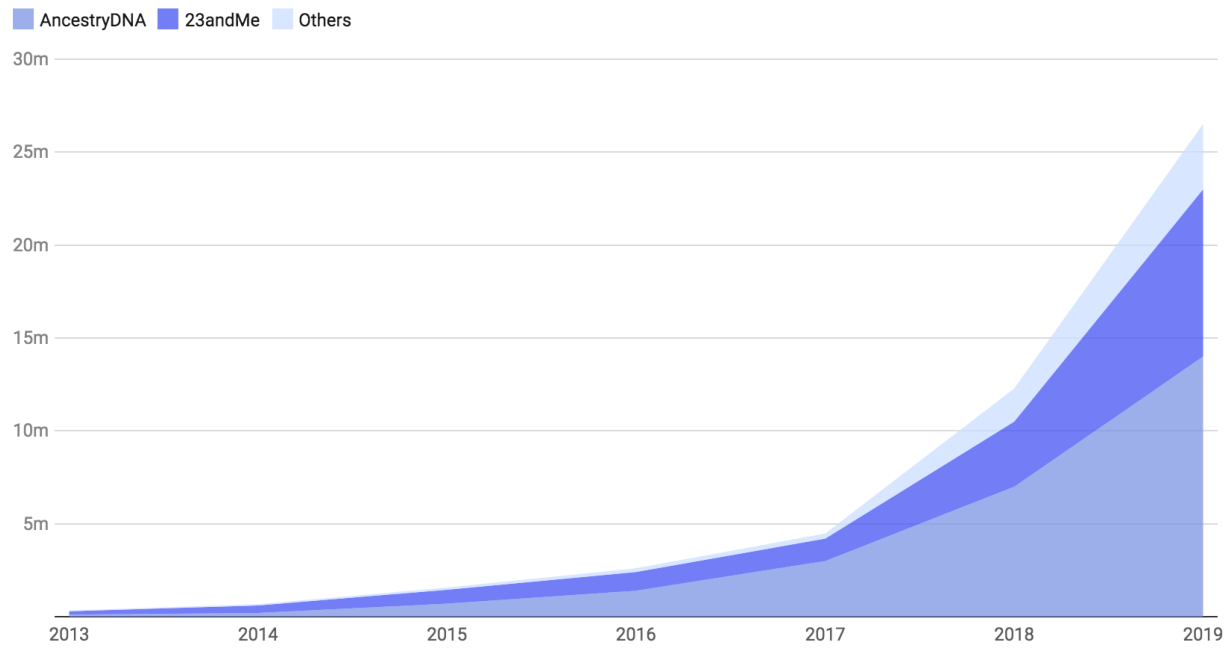


Figure 6. Graph displaying the exponential growth of DTC test use in the past few years. 2017 is a notable turning point, after which the rate of increase becomes significantly greater. Reprinted from "More Than 26 Million People Have Taken an At-Home Ancestry Test," by Regalado, A., 2019, *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/

While about 40 companies exist, Ancestry.com and 23andMe monopolize the commercial genetics industry due to a network effect, meaning that the more individuals who contribute their DNA to one company's database, the more "accurate" the test results become, as DNA from more people and places is compiled for associative analysis (Regalado, 2019; Royal et al., 2010). A couple DTC testing companies market to specific consumers in an effort to offer more precise in-group results or culturally competent services (i.e. African Ancestry). Nonetheless, the network effect exposes an important technical limitation of DTC tests. A news article by Dark Daily cited Adam Rutherford, Ph.D, a British geneticist, expressing that tests are telling

consumers "where on Earth [their] DNA is from today" ("The Problems with Ancestry DNA Analyses," 2018, para. 16) rather than conclusively discerning ancestral *origins*. Not only do they compare DNA samples between consumers who have taken the test to infer ancestral clusters (academic researchers make inferences based on people living today as well), but the diversity of their datasets are subject to those who opt, and have the disposable funds, to purchase kits (Regalado, 2019; Royal et al., 2011). Academic researchers intentionally select which datasets to compare, although using questionable delineations of human populations, while DTC databases are not regulated by a research purpose, meaning that various communities will become, by nature of consumerism, either under- or overrepresented.

Similarly to academic research, the results reported by ancestry tests rely upon the quality of their reference marker panels, which are used to correlate and refine genetic clusters. For example, an alleged new algorithm and reference panel update by Ancestry.com in 2019 spurred discontent among several customers who cited their ethnicity reports changing overnight, significantly shifting some customers' percentages and even erasing various regional affiliations entirely (Pero, 2019). Such flux is a function of the available reference data as well as the chosen statistical approaches and will likely continue to occur as test companies aim to increase precision and database growth. The picture that consumers receive also likely depends upon the research and development underpinning each commercial test. Thinking back to the NDNAD, a forensics database in the UK, several of its unsanctioned auxiliary research projects were a part of a larger field of study surrounding the development of DTC testing; remember, the samples within the NDNAD represent a skewed profile of the UK population, retaining disproportionately high numbers of samples from Black men compared the UK's demographic groups (Wallace, 2011). If DTC test accuracy is a function of its reference database, then

representation is central, although the ability of ancestral inference to accredit geographic origin, and more elusively ethnicity, remains contested regardless.

Beyond their database parameters, some DTC companies utilize panels that functionally differ from academic research. Marker type plays a significant role in the curation of one's ancestral "percentages". In the 2000s, the vast majority of DTC companies employed haploid, or uniparental, mtDNA and Y-chromosome markers, contrasting the use of highly polymorphic autosomal markers in primary research (Royal et al., 2011; Via et al., 2009). Pardo-Seco et al. (2014) attest that mtDNA and Y-chromosome might be functional in certain "genetic contexts" but prove erroneous when applied to inferring "global individual genome ancestry" and can "can only reflect a very tiny portion of the genomic individual ancestry*"* (p.2). In parallel, Brodwin (2002) clarifies a situation that might arise with haploid markers:

> Y chromosomes are passed only through the male line, and an individual has 16 male ancestors in the 5th preceding generation. If you had 1 European ancestor in that generation, and the rest of your male (and female) ancestors were African, then you would be 1/32 European... and… culturally [B]lack in the USA. But if that European man happened to be your father's father's father's father's father, then Y-chromosome typing would place your ancestry entirely in Europe…. (p.328)

Receiving an ancestry report with such skewed results not only provides misleading information but also can be disconcerting to self-concept. While percent breakdowns from DTC tests neither change how an individual navigates the world nor remove the barriers they face, tracing ancestry taps into history; the percentages revealed by genetic genealogy are not numbers devoid of context but can underlie stories of oppression, such as those of colonialism, on a deeply personal level through a medium intrinsically connected to the individual taking the test.

The impact of ancestry tests can be incredibly significant. As a result, it is important that the test-taker understands the limitations, variable accuracy, and probability metrics of DTC tests while evaluating, for one's own self, the role and meaning of these tests in the social context of personal origin. Royal et al. (2011) provides another example of the potential for misinterpretation; they assert that haploid markers, at best, can indicate common ancestry, but leave key questions surrounding the actual origin of that common ancestor unanswered:

> ... if someone lives in North America and [their] mtDNA haplotype exactly matches an individual living in Indonesia, the only thing that can be inferred with confidence is that they share a common ancestor. Without more information about family history and/or the geographic distribution of closely related mtDNA haplotypes, it is impossible to say whether this match arises via recent Indonesia ancestors in North American's family tree, whether both share distant ancestors who lived in entirely different part of the world, or whether the Indonesian match has recent North American heritage. (p. 666)

This implications of this example depend upon how inferences are made by the company's genotyping and visualization technology but theoretically would happen if common ancestry is reported as a percent, or shown as a hotspot, in Indonesia for the North American customer or conversely in North America for the Indonesian customer. A consumer could unknowingly mistake potential common ancestry with another person in the database from a different region to mean that a portion of their ancestry, or "ethnic breakdown", originated from that region. Despite these ambiguities, the breadth and scope of reference data and specific inference techniques behind DTC company tests have remained largely veiled from the public eye, ostensibly in an effort to prevent scientists and consumers alike from dissecting the reliability of the business (Kamala, 2018). As genotyping technology becomes increasingly less costly, more and more

DTC companies might switch to chip-based SNP technology or conduct genome-wide arrays with autosomal markers (Royal et al., 2011). An Ancestry.com blog post announced in 2016 that:

> …we have learned that some markers, also known as SNPs, in DNA are better indicators of ethnic and geographic origins than others, so we have created this new chip to focus on those signals and enable further refinements to the results. This will provide further improvements to the ethnicity results we provide. ("Customer Testing Begins on New AncestryDNA Chip," 2016, para. 2)

This post promotes the idea that SNPs are the key toward more informative "ethnicity" reports. For any given test, it is ultimately impossible to know how accurately the test captures the real human history preceding an individual, and how much the test results for one individual might vary in accuracy from those of another individual; regardless of accuracy, the interpretation of results, influenced partially by company marketing, by concurrent test-takers, and by one's own grasp of genetic ancestry, can maintain real and lasting impacts.

Beyond the nuances of marker type, the timeline of genetic ancestry, specifically among DTC tests, also remains ambiguous, locating one's origins somewhere in between close family and the earliest hominid ancestors (Royal et al., 2011). Duration often plays a significant role in how we perceive our roots; analogously, people will typically distinguish where they were born (especially if they were moved shortly after) from where they grew up. This concept extrapolates to the intersection of genealogy and identity, as individuals could perceive deeper roots in places their ancestors lived for multiple, rather than one, generation or in places where their ancestors voluntarily lived rather than to which they were displaced. Finally, unlike primary literature pools, commercialized tests report on the individual rather than at the population or group level (Royal et al., 2011). Recall that primary research assesses the difference between the

frequencies, or proportions, of gene forms across sample groups rather than the presence of a specific gene form in an individual, and that for almost all genes, any individual in a given human group could possess any form of a particular gene. This fundamental methodological shift imagines genetic interpopulation differences as identifiable in an individual rather than a function of a population, framing "ethnicity", whether intentionally or not, as more discreetly genetic than most human population studies across academia would argue. What does a "breakdown of your ethnicity" (*What to Expect from AncestryDNA®*, n.d., para. 3) mean if ethnicity is a combination of values, institutions, community, and culture, if it is practiced and reinforced through collective memory? Can software updates truly provide more refined insights into ethnicity? Do test reports decontextualize ethnicity and identity? The answers to these questions likely vary, depending upon the cultural identity that a test-taker already holds and their intended purpose for taking the test. Positioning kinship at birth, as Hatton (2019) suggested might lend to conceptualizing ethnicity as proportionally inherited depending on the ethnicities of one's ancestral kin; however, while this version of ethnicity promotes the importance of heritage, it lacks a central interpersonal dimension. Nonetheless, DTC testing is becoming a quintessential part of genealogical searches and identity formation, often ascribed the ability to "discover" or "confirm" one's ancestry.

6.3 Commercialized Genetics, Community, and Personal Origin

Direct-to-consumer tests have expanded the accessibility of genetic technology, exposing the tools of ancestral inference to direct public interaction and transaction. While the majority of consumers choose similar testing formats (i.e. Ancestry.com or 23andMe), experiences interpreting results can vastly differ across individuals and communities. The DTC industry is especially thriving in the US, where people perceive a vast array of branching personal origin

stories (Nelson, 2016). The prevalence of ancestry testing has provoked widespread discussions on the accuracy of genetic ancestry as well as its function in claiming and communicating identity. Nelson argues that, today, "symbolic ethnicity" is prominent among social transactions in America; symbolic ethnicity describes the act of summoning ancestries that are not actively present in one's everyday cultural practices or lived experience, maintaining a symbolic rather than social connection to the implicated ethnicity (Alba, 1990; Nelson, 2016). Instances of symbolic ethnicity include "[harking] back to County Cork, Ireland, while jubilating on St. Patrick's Day in Boston" or "[hanging] an English coat of arms in [one's] house" (Nelson, 2016, p.5). Symbolic ethnicity sustains a selective connection to cultural heritage, whether as distant nostalgia, aesthetic, or social clout. Notably, symbolic ethnicity in America is usually white and Euro-centric, which may relate to the levelling of social, political, and economic positionality of European ethnicities (Alba, 1990) under a expanding concept of whiteness over time, rendering white ethnic distinctions less relevant to power systems.

But, exploration of roots can vary in gravity and impact for those who have experienced ancestral loss or displacement (Nelson, 2016). Brodwin (2002) parallels this sentiment, attesting that genetic ancestry can help tap into a profound sense of connection for persons who "mourn the passage from homeland to diaspora" and "whose collective identity involves the sense of unjust dislocation and culture-loss" (p.327). He proceeds to delve into the differential relationship of genetic ancestry, genealogy, and identity as well as their social and political utility across communities:

> ...certain questions do cry out for anthropological expertise. Why does genetic evidence prove so compelling in some cases (e.g., among diasporic Jews and certain voices in the African American community) and not in others (notably Native Americans)? Why is it

> easily accepted by some groups, but the target of extreme suspicion in others? The availability of genetic tracing surely alters the playing field of identity claims, but it does so differently in each case…. (Brodwin, 2002, p.329)

There is a variable interplay of genes and identity between communities. The impact and utility of genetic ancestry within a community appears to be informed by personal connections to one's origin, the history and heritage underpinning a community, the relationship of the community with hegemonic systems, and the potential for social justice or political mobilization. For some, genetic ancestry offers an unprecedented way to deduce untold stories of diaspora through genealogical branches. Yet, distant and recent pasts of clandestine science targeting marginalized communities simultaneously lends to caution surrounding genetic ancestry.

Genetic ancestry has catalyzed different reactions, understandably spurring distrust while also lending to reconciliation projects and bearing potential for unity, community, and collective memory. One such project is Las Abuelas de Plaza de Mayo, which implemented genetic analysis to connect children with their biological grandparents after they had been nonconsensually displaced into adoptive families throughout post-conflict Argentina (Nelson, 2016). This initiative marks one of the many ways genetic tools have materialized real life outcomes. In her book, Nelson (2016) takes a close look into the "efforts aimed at repairing the social ruptures produced by transatlantic slavery" (p. 9). Nelson describes the significance and nuance of root-finding within African American communities:

> For African Americans, this search is both more elusive and more fraught. A profound loss of social ties was an immediate outcome of the Middle Passage and racial slavery. The ravages of the Civil War left vital records and slave-plantation paperwork degraded or destroyed. (p.5)

In the context of Pan-African identity and the African diaspora, DNA can be a tool for social healing. The potential for DNA to fill in genealogical gaps and trace African ancestry offers a way to discern distant genealogy, cutting through hundreds of years of colonialism with a new layer of clarity. Nelson (2016) also noted that, within the African American community, genetic ancestry has been "annexed onto unresolved and, therefore, persistent debates about national belonging'" (p. 9), proving important for Pan-African social and political agency as well as identity, community, resistance, and pursuit of reparations in post-colonial America. African Ancestry, a DTC test founded by Rick A. Kittles and Gina M. Paige, offers an avenue of deeper connection and "affiliation with nation-states and ethnic groups on the African continent" (Nelson, 2016, p.11). On the company website, African Ancestry offers a search "not to a series of West African REGIONS. But to an ethnic group ('tribe') with specific beliefs, traditions, values and practices" (*African Ancestry PatriClan Test Kit*, n.d., para. 1, emphasis in original). While other companies conflate regional affiliation with ethnicity, Kittle's service markets the ability to trace origin to specific ethnic groups and tribes, not simply regions devoid of cultural heritage. The message of African Ancestry highlights a predominant feature of Ancestry.com and 23andMe, which categorically attest to the ability to reveal "ethnic breakdowns", relaying to consumers a profile of ethnicities (apart from those most salient to their lived experience) they can symbolically reference without necessarily partaking in the culture, values, and barriers of the ethnic communities themselves.

Kittles created African Ancestry with the intention of creating a resource for root-finding that could become a source of healing as well as a unifying social and political force. The idea of providing genetic ancestry for African root-seeking was reported as early as 2000 in the *Los Angeles Times* and proved to be a message that deeply resonated within the African American

community (as cited in Nelson, 2016). Hundreds of individuals proceeded to seek out Kittles while he was a co-director of the molecular genetics unit in the National Human Genome Center at Howard University because they heard he could assist them in reconnecting with and reestablishing "long-lost lineage[s]" (Nelson, 2016, p.11). In celebration of Black History Month during February of 2007, Kittles invited African Americans to the Harlem temple of the Mormon church on Malcolm X Boulevard to share in a root-finding experience, offering free MatriClan and PatriClan (using mtDNA and Y-chromosome DNA respectively) Afrircan Ancestry tests as compensation for research participation; at the time, Dr. Kittles was studying associations between skin pigmentation and various genetic characteristics (Nelson, 2016). Mark Shriver, an author of one of the review articles studied in this thesis, was a co-researcher in this initial study, but after he began adapting it for forensic uses, Kittles stepped away from contributing further, refusing to have a hand in sending more innocent Black people to jail (Nelson, 2016). The event brought together numerous root-seekers, providing a new approach to both individual and community identity formation. Nelson (2016), who was present at the event, describes the sentiments of Kittle's African Ancestry customers, including those she encountered personally:

> ...they spoke of the desire to feel complete, of craving a stronger sense of belonging in the United States and on the continent of Africa, and of wanting in their own way to reckon with the history of slavery…. (p.22)

Ancestral inference offered a powerful elucidation of family lineage, tapping into the pivotal question of "what came before", a discovery process navigating senses of split belonging, reclaiming identity, and forming the ethnic liaisons that exist within Pan-African identity.

DNA can also be politically and legally legitimizing, as seen with Isaiah Washington's access to dual citizenship and assumed sense of duty to Sierra Leone. Nelson (2016) describes

how Kittle's scientific and social mission reaches beyond business; through African Ancestry, he aimed to facilitate goals of racial justice, specifically the "liberation of Black communities through ancestral knowledge" (p. 22). For instance, Kittles' African Ancestry test had a role in a Brooklyn federal court class-action suit in 2002; it was implemented by Deadria Farmer-Paellmann, the attorney as well as founder and director of the Restitution Study Group and the Organization of Tribal Unity in New York, and the plaintiffs to evidence roots in continental Africa and seek restitution for their ancestors' unpaid slave labor and the long-term loss of wealth in the form of both "community and capital" (Nelson, 2016, p.23). While genetic ancestry provides new inlets to both study and actualize ongoing conversations of identity, community, and national belonging within the African American community, there are notable limitations and potential discrepancies beyond the technical accuracy of test kits. Nelson (2016) argues that, while genetic ancestral inference can prove to be "psychically beneficial", DNA results alone neither "materially address persistent structural inequality" nor are "equality, justice, and ethics… easily tethered to or readily settled with DNA evidence" (p.25). Here, Nelson contemplates the intricacies of DNA as a scientific versus social and political tool as well as the utility of DTC ancestry specifically.

The implication of genetic ancestry and identity in legal proceedings opens the door to a plethora of questions as well as restricts evidence of Pan-African identity to scientific validation. Such questions inevitably will speak the language of the testing platform; for example, what is the percent threshold of African ancestry that allows an individual to legally claim compensation for the structural barriers of slavery and discrimination? If no threshold exists, then individuals could attempt to claim marginalized statuses and compensation with only a minute percentage of African ancestry or attempt to reinforce former President Clinton's ideal that "we are all mixed-

race"; however, what does establishing a threshold say about our societal stake in genetic ancestry as well as the authority of genetic ancestry over other sources of personal origin (i.e. oral tradition or memoir)? Additionally, regardless of one's "ethnic percentage", if one navigates the United States as culturally and socially Black, they will face institutional racism and resource inequity nonetheless. In addition to legal frameworks, Isiah Washington's story raises interesting questions surrounding identity and obligation based on DNA. Positioning these questions in conjunction with the fact that the science is still tenuous in multiple facets suggests that DNA could pose promise, but perhaps not in the way it's being popularly mobilized or conceptualized today.

Root-finding can be a momentous exercise for persons who have experienced diaspora and especially significant for those who experienced ancestral loss through the crimes of colonialism or slavery while living in post-colonial nation-states. While Native American communities also faced theft of land and displacement during the colonial period, concepts of personal origin, genealogy, and tribal citizenship do not readily coincide with genetic ancestry. Indigenous communities understand that their interests have rarely been considered by the natural sciences. Louise Erdrich, an author and member of the Turtle Mountain Band of Chippewa Indians, denied an invitation to undergo genetic testing after conversing with tribal elders because the community ultimately "understood her DNA to be communal property" (Nelson, 2016, p.16). While this instance not only relates to conversations surrounding DNA banks and ancestry databases as the pinnacle of big data, which are beyond the scope of this thesis, it also reflects the reality that an individual's DNA can impact the community to which they belong (Nelson, 2016). Furthermore, DTC genetic genealogy operates in a sphere separate

from the scientific, social, and legal frameworks surrounding tribal citizenship and Indigenous identity.

A growing number of Tribes in the United States and First Nations in Canada have been implicating DNA "profiles" or "fingerprints", otherwise known as parentage tests, that verify paternity or close biologically relatives to supplement enrollment applications for tribal citizenship (Tallbear, 2013). After summoning evidence of biological relatedness, an enrollment office might ask for the blood-quantum documentation of a parent to further process for enrollment (Tallbear, 2013). Blood-quantum practices are historically rooted in the concepts and policies surrounding tribal citizenship, described by many scholars as an "incisive social technique for managing Native American lands and peoples" (Tallbear, 2013, p.55). Tallbear (2013) explains that the General Allotment Act of 1887 (or the "Dawes Act") partitioned communal land into individually-owned plots on reservations, only to be inhabited by identifiable Native Americans (Tallbear, 2013). Tallbear cites that the US project endowed Indigenous individuals thought to have more "European blood" or to be "mixed blood" more land because they were perceived as further along evolutionarily trajectory toward civilization and assimilation, the idea being that they would eventually sell off their land, while those deemed more than "half-blood" held 25 year trusts (Tallbear, 2013, p.56); thus, blood quantum sustained a role in dispossessing land through vehemently racist systems. While it is generally affirmed that blood-quantum practices arose due to the imposition of European colonial power systems, codified racism, and ideologies of race and blood, Tallbear also refers to an contrasting perspective:

> [Alexandra] Harmon points out that "in the enrollment councils, federal agents did not brainwash or impose their will on Indians; neither did Indians resolve to draw an

> economically strategic, racially defined boundary around themselves. Rather, officials and Indians participated in a prolonged discourse that I would characterize as incomplete mutual education and accommodation." Harmon refers to the [Northwest Coast Indian enrollment] commissions as "an unprecedented conversation—one that would take place in many tribal communities and continue for decades—about what it meant to be Indian in the twentieth-century United States. (as cited in Tallbear, 2013, p. 53)

Unlike the Euro-centric and colonial-settler lens, such concepts are not supported as biological science, as blood rules preexisted modern conceptualizations of genetic inheritance, among tribal citizens; when talking about full-bloodedness, an interviewee of Jill Doefler, in conjunction to her research on the White Earth Reservation in Minnesota, explained the idea of "full-bloodedness" as a "way of living" rather than a molecular ascription (as cited in Tallbear, 2013, p.52). Tallbear (2013) emphasizes the reality that blood-quantum practices and blood concepts are intangible sociopolitical and legal frameworks, often interpreted as outmoded racialized ideas of purity by both Native Americans and non-Native Americans. Societal and colloquial discourse today oscillate between "semiotic and material meanings of blood and genes" so frequently that many do not realize "blood quantum is a materialist practice only to the extent that it involves paperwork" (Tallbear, 2013, p.54). And while colonialism, and its persisting structures, continue to inform facets of blood-quantum policies, Native American blood concepts do not mold to Euro-centric understandings of molecular relatedness. The language surrounding blood or blood fractions are proxies for a complex cultural and legal scaffolding, as blood quantums not only relay the reservations of one's ancestors but invoke the stories of dispossession and movement of grandparents and their grandparents before (Tallbear, 2013). In parallel, popular conceptualizations of genetic ancestry and genealogy contrast Native American ideas of tribal

belonging and genealogy, as they exist within a scientific framework that continues to be informed by visualizations of molecular personhood:

> The DNA genealogies that are documented by ancestry tests (again, different from the DNA parentage tests that tribes use) and that are co-constituted with hegemonic U.S. race concepts are not yet compatible with the particular biological relationships that tribes privilege. Yet enrollment staff from several tribes told participants at a 2010 national tribal enrollment conference that they had received enrollment applications with commercially purchased genetic-ancestry test results included. This happens even though federally recognized tribes do not accept. (Tallbear, 2013, p.65)

Thus, tribal citizenship, which relies upon both genealogy and established laws, might implicate science in the form of DNA parentage tests; however, tribal citizenship is not compatible with genetic ancestry tests provided by DTC companies. Native American tribes maintain an intricate yet primarily non-molecular understanding of genealogy that encompasses familial relationships as well as "ways of living", cultural practice, and the social processes that tie peoples to both land and community. Discourses surrounding Indigenous identity and tribal citizenship complicate notions of genetic genealogy, rather than the social idea of genealogy, without diminishing the social realities of racialized identities or significance of ancestry within a community.

While the intersection of ancestry and genetics have been manipulated as a vessel of oppression, it has also been reclaimed as tools for reconciliation, root-finding, and restitution. Ancestors who faced oppression as well as oppressors are embedded within genealogies—genealogies carry these stories. It makes sense that the genealogy serves as the archetypal map for navigating both social identity and personal origin. Amidst the era of commercialized

genetics, this map is readily called upon through a scientific lens, often turned to in place of, or to "confirm", historical archive, oral tradition, or familial knowledge; however, the relationship between genes and identity is inherently complicated. Genetic genealogy itself is a social construction, a new method of organizing identity as curated by DTC companies and interpreted by the general public. Companies like Ancestry.com espouse the ability to "connect you to the places in the world where your story started" including to "unique regions," (*Ancestry® Genealogy, Family Trees & Family History Records*, n.d., para. 3) or those considered unique through amidst an ethnocentric perspective. 23andMe relays that its "ancestry breakdown" will help you "Dig deeper into your ancestry, providing the most comprehensive portrait of you yet" (*Ancestry + Traits Service*, n.d., para. 2), or at least on the market. These testaments are promising and enticing, igniting a sense of intrinsic discovery accessible at one's doorstep. On the other hand, scholars like Hatton et al. attest that "to geneticize kinship is an imaginative refigurement of kinship" (Hatton, 2019, p.6), one that boils down oral tradition, written record, or family tradition to seemingly objective valuations. These valuations are molded into profiles of percent breakdowns that make it easy for test takers to quantify an ethnicity (i.e. saying I am "X%" Irish) and qualify it to concurrent origins (i.e. I am 60% Irish but only 2% German).

On a basic level, the interpretability of DTC tests depend upon an understanding of probability and what exactly test results can and cannot tell us about genomic origins. Beyond these inherent limitations, the results also can play a number of different roles in identity formation, sometimes proving to be impactful for those who connect with disrupted roots, casually entertaining for those who unwittingly adopt new ethnic identifiers, and even shocking or confusing for some who receive unexpected results that challenge self-concept. I would argue that identity formation lies far beyond the rigid intervals of an ancestry test or ethnicity

breakdown, specifically as it relates to racial or ethnic identity. Identity informs how an individual navigates the world and experiences kinship. Identity can influence the resource access, discrimination, and systemic barriers an indivdiual faces. Finally, choice is central to identity formation, as it is a social practice involving the building of relationships and community; sometimes, genetic ancestry can be a vessel of community building, but a DTC profile does not dictate an individual's identity. Hauskeller et al. (2013) reviewed various ideas on the social definition of identity, enumerating that recognition by institutions, identity performance, social interactions, power relations and personal choice are key in forming and reinforcing identity (p.877). Receiving unexpected results from a test does not change the way our racial or ethnic identity is perceived or lived, our social and interactional realities, the culture of our childhood and the traditions we practice today, the languages we speak, or the chosen family we have come to know throughout our lives. Personal origin is often a significant facet of identity, as the practice of culture, religion, or other traditions are a way of reproducing the past in the present and connecting with familial or cultural heritage. Origin and ancestry can also impact our social sense of belonging through collective memory or shared history, as Brodwin (2002) describes that "knowledge of ancestry ratifies or even creates a social connection in the present" (p.325). Ancestry can become integrated within the social process of identity formation, but if individuals use genetic ancestry in their toolkit, they should be well informed on what the results can convey to help make individual choices about their meaning in relation to self, family, community, and identity.

CHAPTER SEVEN

Conclusion and Recommendations for Further Research

> In the same way that the earth can be described by many different kinds of maps - from topological to economic - so, too, can the naturally occurring genetic variation among populations be divided in numerous ways and be made to highlight any chosen similarity or difference. (Sankar and Cho, 2002, p.5)

The strategies that geneticists have adapted for ancestral inference are not devoid of ingenuity; they have figured out that highly polymorphic genetic markers arise from benign mutations as well as deduced their microscopic locations for use as molecular locators. Geneticists have developed bioinformatics to visualize the imperceptible, including metrics, such as fixation indices, that have proven reliably functional in other species. Amidst the gaining momentum of genomics and in light of the HGP, they have adapted these techniques to study a detected, yet fleeting, window of human interpopulation variation. Socially informed geneticists like Dr. Rick Kittles, who has critically examined race in genetics and identified the "racial framework" (Batai & Kittles, 2013, p.81) existent in today's biomedical approaches, still continue the search for genetic ancestry, whether for root-finding (i.e. African Ancestry), disease prevention, or other applications. Within his lectures, Kittles aims to make the elusive concept of human variation more accessible:

> So let's say we are looking at a track of DNA:
>
> A C T C A G T T C A
>
> Maybe 94 percent of you guys in the room may have a C at that second-to-last position.
>
> While about 6 percent may have a T:
>
> A C T C A G T T *T* A

That's… a snip… A subtle change. (as cited in Nelson, 2016, p.35)

But, the mystery that surrounds his hypothetical six percent is whether and how it can be explained by individual variability versus population difference. Swynghedauw (2003) pushes us to consider the unconventional approach of an early study:

> More recently, a model-based clustering method… was used to assign individuals to subclusters on the basis of their genotype, ignoring their actual population or racial affiliations (Wilson et al., 2010) … a clustering analysis was carried out to identify four clusters, stopping when an increase in the number of clusters did not enhance the degree of differentiation… Interestingly, 62% of Ethiopians belongs to the same cluster as Norwegians, together with 21% of the Afro-Caribbeans, and the ethnic label 'Asian' inaccurately describe Chinese and Papuans who were placed almost entirely in separate clusters. (p.440)

This study speaks a language that parallels other studies, contending the same fallacies of a racialized framework; however, by employing widely-used techniques and computer programs with a methodological twist, they revealed clusters that intercept the idea of reliably qualifiable human populations. Their approach placed comparatively more limitations on confirmation bias, the notorious crux of pseudo-science.

Genetic ancestry reignites the perceived need to biologically systematize human difference, and geneticists are at the forefront of curating how we see, imagine, and understand human variation, discerning how and whether intergroup differences exist, and constructing the quintessential "human populations" of population genetics. Similarly to how the intricacies of human history can be explored through the evolution of food, linguistics, culture, ideologies, etc., genetics selects the medium of its associations. Geneticists have relied upon identity,

whether regional, ethno-linguistic, religious-cultural, or other social affiliations, to proxy for something far more elusive, something that does not exist within one static timeframe or spatial bounds, something we do not have the vocabulary to effectively describe nor the information to conclusively discern. Substructure, by nature of human movement and generational change, continually restructures. At best, identity can predict, though variably so, these stochastic processes. What would happen if evolutionary events were not conceptualized as predisposed by groups of people but inherited from instances in time? Consider the following hypothetical: if a war were to spur an immense loss of young lives, a bottleneck effect might occur; it is not a social identity but circumstance that is relevant to the gene pool. For all intents and purposes, the gene pool could be defined as all the soldiers across every country involved. And there would be a chance, depending on whether certain alleles were randomly up- or down-regulated within this gene pool, for the next generation descending from the survivors to have a higher likelihood of inheriting certain allele variants.

It is time for genetics to become interdisciplinary. A breadth of perspective is vital not only to navigate the interrelated factors of public health or disease risk but also to recognize how racialization can shape or manifest in their work. Perspective will help geneticists understand their role in conceptualizing human "interpopulation" difference, redesign approach, or even reevaluate the prospective benefit and relative priority of genetic ancestry (as it has been conceptualized thus far) within biomedical or adjunct fields. Geneticists should understand racialization as a socio-historic process so they can recognize continuity in study intent, design, discourse, and application. Furthermore, researchers should neither reduce marginalized communities as unique or convenient opportunities for study nor expose them as targets of disproportionate scrutiny; studies involving these communities must offer full transparency,

ascertain consent, and affirm downstream benefits to the best of their ability. Finally, intentionality is critical regarding the translation of genetics research to the intermediaries of public knowledge. Blell and Hunter (2019) reference a study by Baer et al., which reported how surveyed medical practitioners expressed confusion when trying to discuss race and ethnicity, many "[treating] these concepts as interchangeable and genetically based" (as cited in p.5). Geneticists have a responsibility to prevent the reification of biological race and new forms of biological determinism, racial essentialism, or scientific racism. Systems of oppression have been manipulating scientific theories for hundreds of years because the scientists of those theories have failed to take a decisively anti-racist position.

At the conception of this project, I had aimed to grasp the probability of population genetics and report back what genetics can truly tell us (or not tell us) about identity. My search proved more complex than I could have imagined, opening a floodgate of innumerable branching questions and investigative spheres, pointing toward several opportunities for future research. An examination of media coverage on genetics research could more thoroughly trace the translation of academia to public knowledge. The commercialized era of DTC tests raises pointed questions regarding big data and personal privacy, as one person's genetic information inherently implicates biological relatives (both known and unknown). DTC companies also offer genetic health kits to evaluate disease susceptibility, opening another door to evaluating technical and translational limitations. Both these facets broaden the discussion centering DTC genetic testing. The nascent field of epigenetics is incredibly complex but highly fascinating; a sociological perspective could dissect lines of continuity within the epigenetic realm, deciphering whether ideas of a new "epigenetic determinism" have, or are at risk of, taking root. Finally, geneticists have been making decisions regarding where to look for genetic substructure based on currently-

held concepts of human movement and dispersal through the earliest hominids. These far-reaching theories, which include the termed out-of-Africa model, have been recurrently changed, adapted, dismissed, and reified as various anthropologists, archaeologists, and geneticists have unearthed new fossil evidence. Delving into concurrent theories of human movement, including debates surrounding the social curation of these concepts, would provide an intriguing perspective.

What can genetic ancestry tell us about identity? Elizabeth Warren's misconception extends beyond scientific validity into a line of socio-historic continuity. She invokes an ethnocentric idea of identity as tangible, engaging in the leap between "semiotic and material meanings of blood and genes" (p.54) that Tallbear (2013) describes. She relies upon science as the "final arbiter of truth" (p.4) as termed by Nelson (2016), who explains how society as given intellectual weight and social authority to DNA. Individuals searching for personal origin through the platform of DNA encounter an intensely complex question that not only necessitates a "fairly sophisticated understanding of probability" (Royal et al., 2011, p.668) but also "involves judging the worth of genetic knowledge against other kinds of claims to authentic identity and group membership (oral history, written documentation, cultural practices, inner convictions)" (Brodwin, 2002, p.324). The intersection of DNA with social identity, political utility, and personal origin is an intricate unknown, a social construct that should be approached as such, whereby test-takers and communities have a role in forming its relevance and meaning. The secret of origin is not confined to the materiality of DNA; identity, community, and culture are living concepts, imagined and reimagined, practiced and sustained, challenged and reclaimed, by people.

References

Adams, J. (2008) Obesity, epigenetics, and gene regulation. *Nature Education 1*(1):128. Retrieved from https://www.nature.com/scitable/topicpage/obesity-epigenetics-and-gene-regulation-927/

African Ancestry PatriClan test kit. (n.d.). African Ancestry. Retrieved April 27, 2020, from https://shop.africanancestry.com/products/patriclan-test-kit?variant=4981309997093

Ancestry + traits service. (n.d.). 23andMe. Retrieved April 27, 2020, from https://www.23andme.com/dna-ancestry/?vip=true&pdp=true

Ancestry composition. (n.d.). 23andMe. Retrieved April 27, 2020, from https://medical.23andme.com/wp-content/uploads/2015/10/AncestryComposition.pdf

Ancestry® genealogy, family trees & family history records. (n.d.). Ancestry.Com. Retrieved April 27, 2020, from https://www.ancestry.com/

Alba, R. D. (1990). Conclusion: The emergence of the European Americans. In *Ethnic Identity* (pp. 290–320). Yale University Press; JSTOR. https://www.jstor.org/stable/j.ctt32bp8p.12

Altshuler, D., Donnelly, P., & The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320. https://doi10.1038/nature04226

Astor, M. (2018, October 17). Why many Native Americans are angry with Elizabeth Warren. *New York Times*. Retrieved from https://www.nytimes.com/2018/10/17/us/politics/elizabeth-warren-dna-test.html?module=inline

Azarmandi, M. (2017). Colonial Continuities - A study of anti-racism in Aotearoa New Zealand and Spain (Thesis, Doctor of Philosophy). University of Otago. Retrieved from http://hdl.handle.net/10523/7655

Christina, B. (2010). Teotihuacán. In D. Berger & A. G. Wood (Eds.), *Holiday in Mexico: Critical reflections on tourism and tourist encounters* (pp. 54-76). Duke University Press.

Customer testing begins on new AncestryDNA chip. (2016, May 12). *Ancestry Blog*. Retrieved from https://blogs.ancestry.com/ancestry/2016/05/12/customer-testing-begins-on-new-ancestrydna-chip/

Banihashemi, K. (2009). Iranian human genome project: Overview of a research process among Iranian ethnicities. *Indian Journal of Human Genetics*, *15*(3), 88–92. doi:10.4103/0971-6866.60182

Barker, J. (2004). The human genome diversity project: 'Peoples', 'populations' and the cultural politics of identification *Cultural Studies*, *18*(4), 571–606.

doi:10.1080/0950238042000232244

Batai, K., & Kittles, R. A. (2013). Race, Genetic Ancestry, and Health. *Race and Social Problems*, *5*(2), 81–87. doi:10.1007/s12552-013-9094-x

Bhatia, G., Patterson, N., Pasaniuc, B., Zaitlen, N., Genovese, G., Pollack, S., Mallick, S., Myers, S., Tandon, A., Spencer, C., Palmer, C. D., Adeyemo, A. A., Akylbekova, E. L., Cupples, L. A., Divers, J., Fornage, M., Kao, W. H. L., Lange, L., Li, M., … Price, A. L. (2011). Genome-wide Comparison of African-Ancestry Populations from CARe and Other Cohorts Reveals Signals of Natural Selection. *American Journal of Human Genetics*, *89*(3), 368–381. doi:10.1016/j.ajhg.2011.07.025

Bhopal, R. (2007). The beautiful skull and Blumenbach's errors: The birth of the scientific concept of race. *BMJ : British Medical Journal*, *335*(7633), 1308–1309. doi:10.1136/bmj.39413.463958.80

Blell, M., & Hunter, M. A. (2019). Direct-to-Consumer genetic testing's red herring: "Genetic ancestry" and personalized medicine. *Frontiers in Medicine*, *6*. doi:10.3389/fmed.2019.00048

Blumenbach, J. F., Bendyshe, T., Marx, K. F. H., Flourens, P. (Pierre), Wagner, R., & Hunter, J. (1865). *The anthropological treatises of Johann Friedrich Blumenbach*. London Anthropological Society. Retrieved from http://archive.org/details/anthropologicalt00blumuoft

Boardman, J. D., Daw, J., & Freese, J. (2013). Defining the environment in gene–environment research: Lessons from social epidemiology. *American Journal of Public Health*, *103*(Suppl 1), S64–S72. doi:10.2105/AJPH.2013.301355

Bonham, V. L., Green, E. D., & Pérez-Stable, E. J. (2018). Examining How Race, Ethnicity, and Ancestry Data Are Used in Biomedical Research. *JAMA*, *320*(15), 1533–1534. doi:10.1001/jama.2018.13609

Bowler, P. (2016, May 26). *Social Darwinism*. Oxford Bibliographies. Retrieved from https://www.oxfordbibliographies.com/view/document/obo-9780199756384/obo-9780199756384-0166.xml

Brodwin, P. (2002). Genetics, Identity, and the Anthropology of Essentialism. *Anthropological Quarterly*, *75*(2), 323–330. JSTOR. https://www.jstor.org/stable/3318263

Brunstetter, D. R., & Zartner, D. (2011). Just war against barbarians: revisiting the Valladolid Debates between Sepúlveda and Las Casas. *Political Studies*, *59*(3), 733–752. doi:10.1111/j.1467-9248.2010.00857.x

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, *8*(12). doi:10.1371/journal.pcbi.1002822

Byrd, C., & Hughey, M. (2015). Biological determinism and racial essentialism: The ideological double helix of racial inequality. *The Annals of the American Academy of Political and Social Science*, *661*, 7–22.

Cavalli-Sforza, L. (1994). *The Human Genome Diversity Project*. UNESCO. Retrieved from https://www.osti.gov/servlets/purl/505327

Chial, H. (2008) Rare genetic disorders: learning about genetic disease through gene mapping, SNPs, and microarray data. *Nature Education 1*(1):192

Cooper, R., Rotimi, C., Ataman, S., McGee, D., Osotimehin, B., Kadiri, S., Muna, W., Kingue, S., Fraser, H., Forrester, T., Bennett, F., & Wilks, R. (1997). The prevalence of hypertension in seven populations of west African origin. *American Journal of Public Health*, *87*(2), 160–168.

Costa, J. T. (2009). The Darwinian revelation: Tracing the origin and evolution of an idea. *BioScience*, *59*(10), 886–894. doi:10.1525/bio.2009.59.10.10

Costa, W. P. (2006). European travelers and the writing of the Brazilian nation. In D. H. Doyle & M. A. Pamplona (Eds.), *Nationalism in the New World* (pp. 208–229). University of Georgia Press; JSTOR. https://www.jstor.org/stable/j.ctt46n9z0.17

Crichlow, W. (1993). *The process of "racializaton."* New York: Routledge. Retrieved from http://www.yorku.ca/lfoster/2006-07/sosi4440b/lectures/RACIALIZATION_THEPROCESSOFRACIALIZATION.html

Criollo-Rayo, A. A., Bohórquez, M., Prieto, R., Howarth, K., Culma, C., Carracedo, A., Tomlinson, I., Echeverry de Polnaco, M. M., & Carvajal Carmona, L. G. (2018). Native American gene continuity to the modern admixed population from the Colombian Andes: Implication for biomedical, population and forensic studies. *Forensic Science International: Genetics*, *36*, e1–e7. doi:10.1016/j.fsigen.2018.06.006

Darwin, C. (1871). The descent of man: And selection in relation to sex. London: J.

Darvasi, A., & Shifman, S. (2005). The beauty of admixture. *Nature Genetics*, *37*(2), 118–119. doi:10.1038/ng0205-118

Davis, J. F. (1991). *Who Is Black? One nation's definition*. Pennsylvania State University.

Deng, F. M. (1997, June 1). *Ethnicity: an African predicament.* Retrieved from the Brookings Institution website: https://www.brookings.edu/articles/ethnicity-an-african-predicament/

Dukepoo, F. C. (1999). It's more than the Human Genome Diversity Project. *Politics and the Life Sciences*, *18*(2), 293–297. JSTOR. https://www.jstor.org/stable/4236528

Elhaik, E. (2012). Empirical distributions of FST from large-scale human polymorphism data. *PLOS ONE*, *7*(11), e49837. doi:10.1371/journal.pone.0049837

Fine, M. J., Ibrahim, S. A., & Thomas, S. B. (2005). The role of race and genetics in health disparities research. *American Journal of Public Health*, *95*(12), 2125–2128. doi:10.2105/AJPH.2005.076588

Foster, M. W., & Sharp, R. R. (2002). Race, ethnicity, and genomics: Social classifications as proxies of biological heterogeneity. *Genome Research*, *12*(6), 844–850. doi:10.1101/gr.99202

Frudakis, T. N., & Shriver, M. D. (2004). *Compositions and methods for inferring ancestry*. San Diego, CA. DNA Diagnostics Center Inc, DNAPrint Genomics Inc.

Garrod, J. Z. (2006). A brave old world: An analysis of scientific racism and BiDil®. *McGill Journal of Medicine*, *9*(1), 54–60.

Gayon, J. (2016). From Mendel to epigenetics: History of genetics. *Comptes Rendus Biologies*, *339*(7), 225–230. doi:10.1016/j.crvi.2016.05.009

Gibbs, J. R., & Singleton, A. (2006). Application of Genome-Wide Single Nucleotide Polymorphism Typing: Simple Association and Beyond. *PLOS Genetics*, *2*(10), e150. doi:10.1371/journal.pgen.0020150

Gilson, M., & Tassis, A. (2007, February 1). Genome analysis: Microsatellites or SNPs. *GEN - Genetic Engineering and Biotechnology News*. Retrieved from https://www.genengnews.com/magazine/64/genome-analysis-microsatellites-or-snps/

Graves, J. L. (2003). *The emperor's new clothes: Biological theories of race at the millennium*. Rutgers University Press.

Graves, J. L. (2015). Great is their sin: Biological determinism in the age of genomics. *The ANNALS of the American Academy of Political and Social Science*, *661*(1), 24–50. doi:10.1177/0002716215586558

Grinde, K. E., Brown, L. A., Reiner, A. P., Thornton, T. A., & Browning, S. R. (2019). Genome-wide Significance Thresholds for Admixture Mapping Studies. *American Journal of Human Genetics*, *104*(3), 454–465. doi:10.1016/j.ajhg.2019.01.008

Guess, T. J. (2006). The social construction of whiteness: Racism by intent, racism by consequence. *Critical Sociology*, *32*(4), 649–673. doi:10.1163/156916306779155199

Guidelines for Referring to Populations. (n.d.). Coriell Institute for Medical Research. Retrieved April 25, 2020, from https://www.coriell.org/1/NHGRI/About/Guidelines-for-Referring-to-Populations

Harrell, J. P., Hall, S., & Taliaferro, J. (2003). Physiological responses to racism and discrimination: An assessment of the evidence. *American Journal of Public Health*, *93*(2), 243–248. doi:10.2105/AJPH.93.2.243

Harvey, S. P. (2016). Ideas of race in early america. In *Oxford Research Encyclopedia of American History*. Oxford University Press. doi:10.1093/acrefore/9780199329175.013.262

Hatton, S. B. (2019). History, kinship, identity, and technology: Toward answering the question "what Is (family) genealogy?" *Genealogy*, *3*(1), 2. doi:10.3390/genealogy3010002

Hauskeller, C., Sturdy, S., & Tutton, R. (2013). Genetics and the sociology of identity. *Sociology*, *47*(5), 875–886. doi:10.1177/0038038513505011

Henn, B. M., Gravel, S., Moreno-Estrada, A., Acevedo-Acevedo, S., & Bustamante, C. D. (2010). Fine-scale population structure and the era of next-generation sequencing. *Human Molecular Genetics*, *19*(R2), R221–R226. doi:10.1093/hmg/ddq403

Herndon, A. W. (2018, December 6). Elizabeth Warren stands by DNA test. But around her, worries abound. *New York Times*. Retrieved from https://www.nytimes.com/2018/12/06/us/politics/elizabeth-warren-dna-test-2020.html

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting FST. *Nature Reviews Genetics*, *10*(9), 639–650. doi:10.1126/science.1078311

Howes, R. E., Patil, A. P., Piel, F. B., Nyangiri, O. A., Kabaria, C. W., Gething, P. W., Zimmerman, P. A., Barnadas, C., Beall, C. M., Gebremedhin, A., Ménard, D., Williams, T. N., Weatherall, D. J., & Hay, S. I. (2011). The global distribution of the Duffy blood group. *Nature Communications*, *2*(1), 1–10. doi:10.1038/ncomms1265

Howes, R. E., Battle, K. E., Mendis, K. N., Smith, D. L., Cibulskis, R. E., Baird, J. K., & Hay, S. I. (2016). Global epidemiology of *Plasmodium vivax*. *The American Journal of Tropical Medicine and Hygiene*, *95*(6 Suppl), 15–34. doi:10.4269/ajtmh.16-0141

Human genome sequencing. (2000, June 26). C-SPAN. Retrieved from https://www.c-span.org/video/?157909-1/human-genome-sequencing

International HapMap Consortium. (2004). Integrating ethics and science in the International HapMap Project. *Nature Reviews. Genetics*, *5*(6), 467–475. doi:10.1038/nrg1351

Isern, N., & Fort, J. (2019). Assessing the importance of cultural diffusion in the Bantu spread into southeastern Africa. *PLOS ONE*, *14*(5), e0215573. doi:10.1371/journal.pone.0215573

Jackson, J. P. (2001). "In Ways Unacademical": The Reception of Carleton S. Coon's The Origin of Races. *Journal of the History of Biology*, *34*(2), 247–285. doi:10.1023/A:1010366015968

Florida Society of Health-System Pharmacists (FSHP). (2019, November 25). *John E. Clark receives the ASHP-ABHP Leadership Award*. FHSP News. Retrieved from https://www.fshp.org/news/479492/John-E.-Clark-receives-the-ASHP-ABHP-Leadership-Award.htm

Kahn, J. (2011). BiDil and racialized medicine. In S. Krimsky & K. Sloan (Eds.), *Race and the Genetic Revolution* (pp. 127–141). Columbia University Press; JSTOR. https://www.jstor.org/stable/10.7312/krim15696.12

Kamala, T. (2018, May 15). *Where do Ancestry.com and 23andMe's source datasets come from?* Forbes. Retrieved from https://www.forbes.com/sites/quora/2018/05/15/where-do-ancestry-com-and-23andmes-source-datasets-come-from/#42acda3a52b5

Kliman, R., Sheehy, B. & Schultz, J. (2008) Genetic Drift and Effective Population Size. *Nature Education 1*(3), 3

Lee, S. S.-J. (2015). The biobank as political artifact: The struggle over race in categorizing genetic difference. *The ANNALS of the AAPSS*, *661*(1), 143–159. doi:10.1177/0002716215591141

Lobo, I., & Shaw, K. (2008) Thomas Hunt Morgan, genetic recombination, and gene mapping. *Nature Education 1*(1), 205.

Martinot, S. (2000). The racialized construction of class in the United States. *Social Justice*, *27*(1 (79)), 43–60. JSTOR. https://www.jstor.org/stable/29767190

Masolo, D. A. (2002). Community, identity and the cultural space. *Rue Descartes*, *36*(2), 19–51. doi:10.3917/rdes.036.0019

McWhorter, L. (2009). *Racism and Sexual Oppression in Anglo-America: A Genealogy*. Indiana University Press.

Murji, K., & Solomos, J. (2005). Introduction: Racialization in theory and practice. In K. Murji & J. Solomos (Eds.), *Racialization: Studies in theory and practice* (pp. 1–27). Oxford University Press.

National Human Genome Research Institute (NHGRI). (n.d.). *Polymorphism*. Genome.Gov. Retrieved April 23, 2020, from https://www.genome.gov/genetics-glossary/Polymorphism

Nei, M. (1982). Evolution of human races at the gene level. *Progress in Clinical and Biological Research*, *103 Pt A*, 167–181.

Nelson, A. (2016). *The social life of DNA: Race, reparations, and reconciliation after the genome* (Reprint edition). Beacon Press.

Nievergelt, C. M., Maihofer, A. X., Shekhtman, T., Libiger, O., Wang, X., Kidd, K. K., & Kidd, J. R. (2013). Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel. *Investigative Genetics*, *4*(1), 1–16. doi:10.1186/2041-2223-4-13

Nogueira, S. G. (2013). Ideology of white racial supremacy: Colonization and de-colonization processes. *Psicologia & Sociedade*, *25*(SPE), 23–32. doi:10.1590/S0102-71822013000500004

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. doi:10.1016/j.pcl.2015.05.012

Novembre, J., & Peter, B. M. (2016). Recent advances in the study of fine-scale population structure in humans. *Current Opinion in Genetics & Development*, *41*, 98–105. doi:10.1016/j.gde.2016.08.007

Pääbo, S. (2003). The mosaic that is our genome. *Nature*, *421*(6921), 409–412. doi:10.1038/nature01400

Papadopoulos, D. (2011). The Imaginary of Plasticity: Neural Embodiment, Epigenetics and Ecomorphs. *The Sociological Review*, *59*(3), 432–456. doi:10.1111/j.1467-954X.2011.02025.x

Pardo-Seco, J., Martinón-Torres, F., & Salas, A. (2014). Evaluating the accuracy of AIM panels at quantifying genome ancestry. *BMC Genomics*, *15*(1), 1–13. doi:10.1186/1471-2164-15-543

Pereira, V., Freire-Aradas, A., Ballard, D., Børsting, C., Diez, V., Pruszkowska-Przybylska, P., Ribeiro, J., Achakzai, N. M., Aliferi, A., Bulbul, O., Carceles, M. D. P., Triki-Fendri, S., Rebai, A., Court, D. S., Morling, N., Lareu, M. V., Carracedo, Á., & Phillips, C. (2019). Development and validation of the

EUROFORGEN NAME (North African and Middle Eastern) ancestry panel. *Forensic Science International: Genetics*, *42*, 260–267. doi:10.1016/j.fsigen.2019.06.010

Pero, J. (2019, April 29). Ancestry.com users upset after database update alters ethnicities. *DailyMail*. Retrieved from https://www.dailymail.co.uk/sciencetech/article-6972711/Ancestry-com-fire-update-database-drastically-changes-ethnicity-users.html

Portela, A., & Esteller, M. (2010). Epigenetic modifications and human disease. *Nature Biotechnology*, *28*(10), 1057–1068. doi:10.1038/nbt.1685

Prather, C., Fuller, T. R., Jeffries, W. L., Marshall, K. J., Howell, A. V., Belyue-Umole, A., & King, W. (2018). Racism, African American women, and their sexual and reproductive health: A review of historical and contemporary evidence and implications for health equity. *Health Equity*, *2*(1), 249–259. doi:10.1089/heq.2017.0045

Regalado, A. (2019, February 11). More than 26 million people have taken an at-home ancestry test. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/

Reill, P. H. (2005). *Vitalizing nature in the Enlightenment*. University of California Press.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, *298*(5602), 2381–2385. doi:10.1126/science.1078311

Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of Genetic Markers for Inference of Ancestry*. *The American Journal of Human Genetics*, *73*(6), 1402–1422. doi:10.1086/380416

Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J., & Clark, A. G. (2010). Inferring genetic ancestry: opportunities, challenges, and implications. *American Journal of Human Genetics*, *86*(5), 661–673. doi:10.1016/j.ajhg.2010.03.011

Ryan, H. (2019, May 28). How eugenics gave rise to modern homophobia. *The Washington Post*. Retrieved from https://www.washingtonpost.com/outlook/2019/05/28/how-eugenics-gave-rise-modern-homophobia/

Saeb, A. T. M., & Al-Naqeb, D. (2016). The impact of evolutionary driving forces on human complex diseases: A population genetics approach. *Scientifica*. doi:10.1155/2016/2079704

Salzano, F. M., & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology*, *37*(1 Suppl), 151–170.

Sankar, P., & Cho, M. K. (2002). Toward a new vocabulary of human genetic variation. *Science (New York, N.Y.)*, *298*(5597), 1337–1338. doi:10.1126/science.1074447

Sawyer, P. J., Major, B., Casad, B. J., Townsend, S. S. M., & Mendes, W. B. (2012). Discrimination and the Stress Response: Psychological and Physiological Consequences of Anticipating Prejudice in Interethnic Interactions. *American Journal of Public Health*, *102*(5), 1020–1026. doi:10.2105/AJPH.2011.300620

Shi, W., Ayub, Q., Vermeulen, M., Shao, R., Zuniga, S., van der Gaag, K., de Knijff, P., Kayser, M., Xue, Y., & Tyler-Smith, C. (2010). A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP–CEPH populations. *Molecular Biology and Evolution*, *27*(2), 385–393. doi:10.1093/molbev/msp243

Short, A. (2016, February 14). Bill Clinton downplays Obama: "We're all mixed-race people." *New York Post*. Retrieved from https://nypost.com/2016/02/14/bill-clinton-downplays-obama-were-all-mixed-race-people/

Shreffler, K. M., McQuillan, J., Greil, A. L., & Johnson, D. R. (2015). Surgical sterilization, regret, and race: Contemporary patterns. *Social Science Research*, *50*, 31–45. doi:10.1016/j.ssresearch.2014.10.010

Shriner, D. (2013). Overview of admixture mapping. *Current Protocols in Human Genetics*, *76*(1), 1.23.1-1.23.8. doi:10.1002/0471142905.hg0123s76

Shufeldt, R. W. (1915). *America's greatest problem: The Negro*. Philadelphia, F. A. Davis. Retrieved from http://archive.org/details/americasgreatest00shuf

Simmons, D. (2008) Epigenetic influence and disease. *Nature Education* 1(1):6. Retrieved from https://www-nature-com.duproxy.palni.edu/scitable/topicpage/epigenetic-influences-and-disease-895/

Simón, Y. (2018). Hispanic vs. Latino vs. Latinx: A brief history of how these words originated. *Remezcla*. Retrieved from https://remezcla.com/features/culture/latino-vs-hispanic-vs-latinx-how-these-words-originated/

Skwiot, C. (2010). Garden republics or plantation regimes? In *The Purposes of Paradise* (pp. 49–86). University of Pennsylvania Press; JSTOR. https://www.jstor.org/stable/j.ctt3fhf6t.5

Smith, M. W., & O'Brien, S. J. (2005). Mapping by admixture linkage disequilibrium: Advances, limitations and guidelines. *Nature Reviews Genetics*, *6*(8), 623–632. doi:10.1038/nrg1657

Steers-Mccrum, A. R. (2018). Out of the binary and beyond the spectrum: Redefining and reclaiming Native American race. *Critical Philosophy of Race*, *6*(2), 216. doi:10.5325/critphilrace.6.2.0216

Stepan, N. (2001). *Picturing Tropical Nature*. Cornell University Press.

Stern, A. M. (2016). Eugenics, sterilization, and historical memory in the United States. *História, Ciências, Saúde-Manguinhos*, *23*, 195–212. doi:10.1590/s0104-59702016000500011

Strmic-Pawl, H. V., Jackson, B. A., & Garner, S. (2018). Race Counts: Racial and Ethnic Data on the U.S. Census and the Implications for Tracking Inequality. *Sociology of Race and Ethnicity*, *4*(1), 1–13. doi:10.1177/2332649217742869

Swynghedauw, B. (2003). Human races and evolutionary medicine. *European Review; Cambridge*, *11*(3), 437–447. http://dx.doi.org.duproxy.palni.edu/10.1017/S1062798703000371

TallBear, K. (2013). *Native American DNA: Tribal belonging and the false promise of genetic science*. University of Minnesota Press.

Takezawa, Y. (2012). Problems with the terms: "Caucasoid", "Mongoloid" and Negroid". *Kyoto University Research Information Repository*, *43*, 61–68. doi:10.14989/155688

The Indian Genome Variation Consortium. (2005). The Indian Genome Variation database (IGVdb): A project overview. *Human Genetics*, *118*(1), 1–11. doi:10.1007/s00439-005-0009-9

The problems with ancestry DNA analyses. (2018, October 12). *DARK Daily Laboratory and Pathology News*. Retrieved from https://www.darkdaily.com/the-problems-with-ancestry-dna-analyses/

Vartanian, A. (1953). *Diderot and Descartes: A study of scientific naturalism in the Enlightenment*. Princeton University Press.

Verdu, P., Pemberton, T. J., Laurent, R., Kemp, B. M., Gonzalez-Oliver, A., Gorodezky, C., Hughes, C. E., Shattuck, M. R., Petzelt, B., Mitchell, J., Harry, H., William, T., Worl, R., Cybulski, J. S., Rosenberg, N. A., & Malhi, R. S. (2014). Patterns of admixture and population structure in Native populations of northwest North America. *PLOS Genetics*, *10*(8), e1004530. doi:10.1371/journal.pgen.1004530

Via, M., Ziv, E., & Burchard, E. G. (2009). Recent advances of genetic ancestry testing in biomedical research and direct to consumer testing. *Clinical Genetics*, *76*(3), 225–235. doi:10.1111/j.1399-0004.2009.01263.x

Viola, L., & Verheul, J. (2019). The Media Construction of Italian Identity: A Transatlantic, Digital Humanities Analysis of *italianità* , Ethnicity, and Whiteness, 1867-1920. *Identity*, *19*(4), 294–312. doi:10.1080/15283488.2019.1681271

Von Essen, S. G., & Auvermann, B. W. (2005). Health effects from breathing air near CAFOs for feeder cattle or hogs. *Journal of Agromedicine*, *10*(4), 55–64. doi:10.1300/J096v10n04_08

von Humboldt, A. (1995). *Personal narrative of a journey to the equinoctial regions of the new continent* (J. Wilson, Trans.). Penguin Classics.

Walker, R. J., Williams, J. S., & Egede, L. E. (2016). Impact of race/ethnicity and social determinants of health on diabetes outcomes. *The American Journal of the Medical Sciences*, *351*(4), 366–373. doi:10.1016/j.amjms.2016.01.008

Wallace, H. (2011). Prejudice, stigma, and dna databases. In S. Krimsky & K. Sloan (Eds.), *Race and the Genetic Revolution* (pp. 68–96). Columbia University Press; JSTOR. https://www.jstor.org/stable/10.7312/krim15696.9

Westerlund, J. F., & Fairbanks, D. J. (2010). Gregor Mendel's classic paper and the nature of science in genetics courses. *Hereditas*, *147*(6), 293–303.

What to Expect from AncestryDNA®. (n.d.). Ancestry.com. Retrieved April 16, 2020, from https://support.ancestry.com/s/article/What-to-Expect-from-AncestryDNA

Wood, J., & Chasteen, J. (2009). *Problems in Modern Latin American History: Sources and Interpretations* (Third). Rowman and Littlefield.

Yudell, M. (2011). A short history of the race concept. In S. Krimsky & K. Sloan (Eds.), *Race and the Genetic Revolution* (pp. 13–30). Columbia University Press; JSTOR. https://www.jstor.org/stable/10.7312/krim15696.6

Zenou, Y., & Boccard, N. (2000). Racial discrimination and redlining in cities. *Journal of Urban Economics*, *48*(2), 260–285. doi:10.1006/juec.1999.2166

Zheng-Bradley, X., & Flicek, P. (2017). Applications of the 1000 Genomes Project resources. *Briefings in Functional Genomics*, *16*(3), 163–170. doi:10.1093/bfgp/elw027

Zhu, X., Young, J. H., Fox, E., Keating, B. J., Franceschini, N., Kang, S., Tayo, B., Adeyemo, A., Sun, Y. V., Li, Y., Morrison, A., Newton-Cheh, C., Liu, K., Ganesh, S. K., Kutlar, A., Vasan, R. S., Dreisbach, A., Wyatt, S., Polak, J., … Levy, D. (2011). Combined admixture mapping and association analysis identifies a novel blood pressure genetic locus on 5p13: Contributions from the CARe consortium. *Human Molecular Genetics*, *20*(11), 2285–2295. doi:10.1093/hmg/ddr113