A Reconsideration of the Role of Self-Identified Races in Epidemiology and Biomedical Research

[Postprint. Please cite version in Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, Special issue: "Genomics and Philosophy of Race", 52, 2015, pp. 56-64.]*

A Reconsideration of the Role of Self-Identified Races in Epidemiology and Biomedical Research

Ludovica Lorusso* and Fabio Bacchini**

* Department of Political Science, Communication, Engineering and Information Technologies, University of Sassari, Italy (corresponding author: Viale Mancini 5, 07100 Sassari, Italy. lorusso@uniss.it)
** Laboratory of Applied Epistemology, DADU, University of Sassari, Italy

Abstract

A considerable number of studies in epidemiology and biomedicine investigate the etiology of complex diseases by considering (self-identified) race as a relevant variable and focusing on the differences in risk among racial groups in the United States; they extensively draw on a genetic hypothesis - viz. the hypothesis that differences in the risk of complex diseases among racial groups are largely due to genetic differences covarying with genetic ancestry - that appears highly problematic in the light of both current biological evidence and the theory of human genome evolution. Is this reason for dismissing self-identified races? No. An alternative promising use of self-identified races exists, and ironically is suggested by those studies that investigate the etiology of complex diseases without focusing on racial differences. These studies provide a large amount of empirical evidence supporting the primacy of the contribution of non-genetic as opposed to genetic factors to the risk of complex diseases. We show that differences in race – or, better, in racial self-identification – may be critically used as proxies for differences in risk-related exposomes and epigenomes in the context of the United States. Self-identified race is what we need to capture the complexity of the effects of present and past racism on people's health and investigate risk-related external and internal exposures, gene-environment interactions, and epigenetic events. In fact patterns of racial self-identifications on one side, and patterns of risk-related exposomes and epigenomes on the other side, constantly coevolve and tend to match each other. However, there is no guarantee that using self-identified races in epidemiology and biomedical research will be beneficial all things considered: special attention must be paid at balancing positive and negative consequences.

1. *Introduction*

In the contemporary United States, the risk of morbidity and mortality from most complex or multifactorial disease is patterned along racial lines. For example, in 2009, the non-Hispanic black population had 141.3 deaths per 100,000 (age adjusted) that were due to coronary heart disease, compared with 117.7 deaths per 100,000 in the non-Hispanic white population.[1] In the same year the non-Hispanic white population had 37.8 deaths per 100,000 that were due to stroke, compared to 55.7 deaths per 100,000 in the non-Hispanic black population; non-Hispanic black females had 31.2 deaths per 100,000 population (age adjusted) that were due to breast cancer, over two and a half times the rate among Asian or Pacific Islander females, 11.4 per 100,000 (Healthy People 2014).

The main aim of this article is to critically examine whether these differences are important in the study of the etiology of complex diseases - that is, whether "race" (specifically, self-identified race) can be usefully employed as a variable in the research into the causes of the susceptibility to complex diseases in the human

[1] According to the U.S. Office of Management and Budget (OMB), races in the U.S. are Black, White, Asian, American Indian, and Pacific Islander. "Hispanic" and "non-Hispanic" are ethnicities, not races.

population.

Indeed a distinction can be drawn in epidemiology and biomedical research among two categories of studies, which we call, respectively, "race-based studies" (RBS) and "race-neutral studies" (RNS). We define RBS as those epidemiological and biomedical studies investigating the etiology of complex diseases which do employ race as a relevant variable in their study design, thus focusing (among other things) on racial differences in the disease risks in the search for determinants of disease; as a consequence, these studies assume race to be a proxy for some causal factors on the pathway leading to disease which can either be specified or remain unspecified. On the other hand, we define RNS as those epidemiological and biomedical studies investigating the etiology of complex diseases which do not employ race as a relevant variable, do not focus on racial differences in the disease risks in the search for determinants of disease, and therefore avoid considering race as a proxy for any causal factors on the pathway leading to disease. While RBS consider it important to direct attention to associations between (self-identified) race and complex disease phenotypes, one of their goals being to determine a "racial" susceptibility to complex diseases, RNS do not consider such associations as relevant and only aim to determine a susceptibility to complex diseases which is not supposed to be race-specific.

This distinction clearly emerges in the literature, where many authors have similarly opposed "race-based" to "race-neutral" research, especially in genetic epidemiology and biomedicine - arguably because the use of race as a proxy for genetic features is considered ethically problematic and scientifically controversial. For instance, Fujimura & Rajagopalan (2011) remarks that in the field of biomedical genetic research it is possible to distinguish among those scientists who employ the race variable in the investigation of the etiology of complex diseases and those who think that "race categories [...] are not appropriate tools to search for disease-related genes" (p. 6). Paradies et al. (2007) differentiate between "race-neutral approaches" and approaches acknowledging race as a proxy of either genetic or social and environmental factors. In examining the use of racial categories deployed to explain specific disease patterns, Fausto-Sterling (2008) explicitly contrasts "medical scientists continu[ing] to study racial differences" (p. 659), "addressing racial differences" in disease study (p. 661), and assuming that "race might be an important study variable" (p. 662), to "race-neutral approaches" (p. 666). Similarly, Baer et al. (2013) oppose "health researchers consider[ing] race and ethnicity useful categories for health research" (p. 212) to those avoiding "the concept of race as a useful unit of analysis" (p. 213). Evidently the distinction between RBS and RNS is not new and has emerged various times in the relevant literature, however in slightly different forms (see also e.g. Shields et al. 2005; Fujimura et al. 2008).

Notice that the distinction we introduce between RBS and RNS is orthogonal to the distinction among those studies that stress the role of genetic factors, and those ones that stress the role of non-genetic factors, in the explanation of the risks of complex diseases. Although assigning importance to information on race in the investigation of the etiology of complex diseases does not necessarily require favoring the causal role of genetic factors, a noticeable number of RBS in epidemiology and biomedical research[2] assume (however sometimes implicitly) that the race variable can play a relevant role as a proxy for genetic causal factors – as opposed to non-genetic causal factors – importantly contributing to the risk of complex diseases (Rebbeck et al. 2006; Frank 2007; Paradies et al. 2007; Lee 2009; Megyesi et al. 2011).

In particular, RBS extensively adopt what we call the *genetic hypothesis*: they assume that differences in the risk of complex diseases among racial groups are largely due to genetic differences covarying with genetic ancestry which self-identified races are supposed to be good proxies for (e.g. Burchard et al. 2003; Kistka et al. 2007; Drake et al. 2008; Campbell & Tishkoff 2008; Eeles et al. 2014; Levin et al. 2014). Such genetic differences may consist in differences either in population-specific genetic variants or in genetic variants differentially distributed among populations (e.g. Kumar et al. 2010; Aldrich et al. 2012). Thus self-identified races can be used as proxies for a population-specific genetic component to the risk of complex diseases.

In this paper we argue that the genetic hypothesis is not immune from decisive criticism, and show that RBS seem up a blind alley in their use of self-identified races as proxies for genetic risk variants. However, this does not mean – as one may expect – that as a consequence self-identified races have no useful role to play in the research into the etiology of complex diseases. An alternative and epistemologically correct use of self-identified races exists, and ironically is suggested by those very studies that do not focus on races and make no use of self-identified races – viz. RNS.

In fact RNS provide a large amount of empirical evidence supporting the primacy of the contribution of a

[2] Note, however, that in epidemiology the use of (self-identified) race as a proxy for risk-related genetic factors seems less common than it is in biomedical research.

non-genetic as opposed to a genetic kind of variation to the risk of complex diseases (e.g. Vineis et al. 2009; Rappaport & Smith 2010; Miller & Jones 2014). We show, then, that self-identified races may be critically used as proxies for a risk-related environmental and epigenetic variation in the context of the United States. Self-identified race is what we need to capture the complexity of the effects of present and past racism on people's health and investigate risk-related external and internal exposures, gene-environment interactions, and epigenetic events. Our point is that a promising category of studies into the etiology of complex diseases is that of RBS focusing on non-genetic causal factors: self-identified race can be correctly employed as a useful variable in epidemiology and biomedical research, provided that the genetic hypothesis is dismissed.

2. *The genetic hypothesis in RBS*

In the genetic hypothesis self-identified races are considered as proxies for a specific genetic ancestry associated with specific genetic variants contributing to the risk of complex diseases (e.g. Kumar et al. 2010; Bustamante et al. 2011; Fejerman et al. 2013). Generally genetic ancestry itself is considered unlikely to be the cause of the population-specific genetic susceptibility, but is taken as a proxy for genetic variants contributing to the risk, which are supposed to be either population-specific or differentially distributed among populations (Aldrich et al. 2012). So the genetic hypothesis is based on three main assumptions:

1. Self-identified race is a good proxy for a specific genetic ancestry.
2. Different specific genetic ancestries can be correctly and unambiguously identified.
3. A specific genetic ancestry can be used as a proxy for unknown genetic variants contributing to the risk of a complex disease, which are supposed to be either population-specific or differentially distributed among populations.

These three assumptions are problematic for different reasons. The first assumption is problematic because racial self-identification depends upon several kinds of psychological, cultural, and social factors - a fact that makes it complex and ambiguous, and therefore not a robust proxy for genetic ancestry (Hunt & Megyesi 2008). For instance, racial self-identification clearly depends upon personal opinions about such issues like what races are, how central racial identity should be in people's lives, and the like.[3] Furthermore, it is contingent upon the particular list of admissible races to be chosen from (e.g. Saperstein & Penner 2012). Individuals assumed to share a specific genetic ancestry might easily differ in their racial self-identification across different historical and sociocultural contexts (e.g. Rotimi & Jorde 2010), and even across dissimilar socioeconomic status, neighborhood or perceived racial discrimination (see below, section 4). Conversely, individuals who share racial self-identification may turn out to be attributed very different genetic ancestries (e.g. Bryc et al. 2010). Therefore, the accuracy in the prediction of genetic ancestry from racial self-identification is very low, especially in the case of the so-called "racially admixed" populations, which are characterized by complex recent ancestral histories.

The second assumption concerns the possibility itself of correctly and unambiguously identifying different specific genetic ancestries. Consider in particular the case of admixed populations, the most used in RBS. In order to identify the proportion of different specific genetic ancestries in the individuals of these populations, geneticists (e.g. Fejerman et al. 2009) usually rely on different software based on specific statistical approaches to estimate ancestry (e.g. Bayesian methods or methods based on maximum-likelihood estimation), which treat our species in admixture terms, as "if populated by people who either are members of discrete populations or are admixed descendants of such populations" (Weiss & Long 2009, p. 704). The assumption of the existence of some prior, pure, and distinct ancestral populations, however, is in contrast with the evolutionary history of the human genome, according to which "each genome [is] a mosaic of haplotype blocks, each with its own origin and history, brought together in the same cell by sexual reproduction" (Barbujani et al. 2013, p. 157). In addition, when you want to search for percentages of genetic ancestry in a particular individual, you need to compare her DNA variation with that of samples coming from populations *presumed* to be proxies for her ancestral populations. However, the choice of the proxy populations is at least partially arbitrary (e.g. Salas et al. 2005), therefore each attribution of a specific genetic ancestry to an individual is in itself ambiguous (e.g. Royal et al. 2010). Also consider that clustering methods employed by population geneticists to assign individual organisms to distinct statistical clusters using genetic data are model-based, thus the significance of such attributions must be always considered as dependent on both the specific model assumed by the clustering

[3] For the importance of biology education in shaping these personal opinions, see Donovan (forthcoming).

method and the data provided (e.g. Kaplan & Winther 2013; Winther et al. forthcoming).

The third assumption concerns the possibility of inferring a certain risk-associated genotype from a certain identified genetic ancestry. This assumption seems highly questionable if interpreted as the assumption that a robust and unambiguous genetic clustering of human populations exists, according to which individuals of each population are characterized either by similar frequencies in genetic variants that are differentially distributed among populations, or by shared population-specific variants. This way of interpreting the assumption is in fact not supported by evidence from the history of human evolution, according to which gene flow through migrations has been the main evolutionary force making up human genome diversity; as a consequence genetic traits are extensively shared in humans across the world (e.g. Jakobsson et al. 2008; Rotimi & Jorde 2010) - they tend to be continuous across human populations, and are not generally correlated in their variation (e.g. Wilson et al. 2001; Barbujani 2005). Even if it is possible to use differences in frequency of some genetic traits to structure the human species along continental lines (e.g. Rosenberg et al. 2002; Li et al. 2008)[4], we hardly can use any genetic structure to infer the variation of other traits of the individuals' genome. In other words, there is no guarantee that any particular genetic structure of human beings - built through a specific methodology and by using specific traits - can be used to predict variation in other traits (Barbujani et al. 2013).

However, if differently interpreted, the third assumption can make a biologic sense. In fact, if we analyze a large number of alleles of an individual's genome, and the majority of these alleles turn out to be more frequent in - say - African populations, given an unknown genetic variant that is supposed to be more frequent in (or, specific to) African populations, it is probable that the individual will show that particular variant. Therefore, by analyzing a large portion of genome variation of an individual, we may infer with a good statistical confidence the presence of other variants. Nonetheless, for the third assumption to stand it is not sufficient that we can make this kind of inference; it is also necessary that the inferred genetic variant that is supposed to be either population-specific or differentially distributed among human populations be *also a risk-related variant*.

Consider that the genetic variants involved in determining a susceptibility to a complex disease can be either rare or common. Rare genetic variants (frequency < 3%) are supposed to be more likely the ones that may either differ in frequency between different human populations or even be population-specific. In fact, rare variants are those variants that have arisen more recently in the history of the human species, therefore they are more likely than common variants to be differentially distributed among populations of different continents (e.g. Manolio et al. 2009). However, only few rare variants have been found to be associated to complex diseases, and recent analyses of the allele frequency distributions of data from genome-wide association studies (GWAS) strongly limit the number of rare variants and the range of their effect sizes that would be compatible with them making a large contribution to disease risk (Wray et al. 2011; for a review of the arguments against a major role of rare alleles in the explanation of the risk of complex diseases see Gibson 2012).

The second possibility is that the genetic variants contributing to the risk of a complex disease are common variants. In fact, some researchers pushing the genetic line have suggested that differences in the risk of complex diseases among racial groups might be due to differences in the additive effects of a large number of common allelic variants of small effect. However, common alleles are typically widely distributed across populations, only seldom show high levels of linkage disequilibrium with other genome markers, and hence their value in predicting a risk of complex disease is limited. As a consequence, genetic ancestry is not likely to be a good proxy for this kind of genetic component to the risk. Furthermore, as we will see in the next section, additive genetic effects appear able to predict a very small amount of the risk of complex diseases in the general population. Thus it seems very questionable to postulate that differences in additive genetic effects can importantly contribute to explanations of differences in the risk of complex diseases among racial groups in the first place.

We can conclude that also the third assumption is considerably problematic; and, since the first two assumptions above appear to be supported by neither theoretical considerations nor empirical evidence in human genetics, the genetic hypothesis as a whole seems seriously undermined.

3. *The study of complex diseases in RNS*

[4] It is worth noting that according to some race scholars human population structure results provide evidence for biological realism about racial groups in contemporary ordinary racial discourse in the U.S. See e.g. Spencer (forthcoming).

How have the risks of complex diseases been investigated so far in RNS? Researchers have studied risk-associated genetic and environmental factors and their role in determining the development of the diseases within the general population (e.g. Hung et al. 2008; Hindorff et al. 2009; Thomas 2010; Patel et al. 2010; Gibson 2012). It is possible to individuate two main families of rival explanatory models of the risk of a complex disease in RNS which differ in the emphasis they put on the environmental factors.

The first family of models ('gene-based' models) mainly focus on the causal role of genetic variants, which determine an *individual genetic susceptibility* to the disease; these variants are supposed to cause the development of complex diseases in the presence of specific environmental exposures that participate with different modalities in the processes leading to disease (e.g. Hung et al. 2008; Antoniou et al. 2014). While the main actor is the genotype and the biological mechanisms determined by genetic factors, the role of the environment is providing a context that acts as a modifier of their power and is thought to be separable from them. A gene-based model clearly also underlies the genetic hypothesis in RBS, where differences in the risks of complex diseases among racial populations are mostly due to differences in a *population-specific genetic susceptibility*.

The second family of models ('environment-based' models), on the other hand, mainly focus on the causal role of environmental exposures (here considered in the broadest sense and including, for instance, lifestyles, culture-dependent behaviors, social interactions, etc. - see e.g. Miller & Jones 2014), gene-environment interactions (e.g. Thomas 2010), and epigenetics (e.g. Gluckman et al. 2010). The role of individual genetic susceptibility is backstage and is only supposed to play a major role in modulating individual risk when environmental exposures' levels are low (e.g. Vineis et al. 2009). The family of environment-based models is not just the symmetric counterpart of the family of gene-based models: since models of this family posit that the environmental exposures can become embodied in the biologic systems and produce an *acquired susceptibility* to a complex disease (e.g. via epigenetic changes that may either occur during a person's life or being inherited), they may be said to go beyond the nature-nurture clear-cut division itself.

While in different periods RNS have fluctuated between the two families of models, after the start of the Human Genome Project in the 2000s they mainly focused on the gene-based models, assuming that common complex diseases are mainly due to the additive effects of common genetic variants of low penetrance. A huge number of RNS systematically searched for associations between common variants or SNPs (Single Nucleotide Polymorphisms) across the whole genome and complex phenotypic traits or diseases; the SNPs found to be associated to traits or diseases have been called TASs, trait/disease-associated SNPs (Hindorff et al. 2009). Thousands of TASs have been identified so far, but the impact of GWAS is, however, quite disappointing, given that these associations are able to predict a very small amount of the risk of complex diseases in individuals; this problem has been called the "missing heritability" problem (Manolio et al. 2009).[5] In recent years many doubts have arisen about GWAS and the use of TASs in preventive medicine: "An important question is to what extent GWAS have identified genetic variants likely to be of clinical or public health importance, particularly for developing preventive or therapeutic interventions" (Hindorff et al. 2009, p. 9366). In fact very few TASs have been shown to be the actual risk variants, and a plethora of them could not be replicated in independent studies, thus were considered to be spurious (Chanock et al. 2007).

Searching for an explanation of the missing heritability, RNS recently reconsidered the role of rare genetic variants of large effect as well as of non-additive effects of combinations of genetic variants, like dominance[6] and epistasis[7], in genetic variance (e.g. Shao et al. 2008; Gibson 2012); more importantly, many studies started to focus deeply on non-genetic variance, viz. environmental factors, gene-environment interactions, and epigenetics, thereby exploring the family of the environment-based models (Vineis et al. 2009; Eichler et al. 2010; Feinberg 2007, 2010; Rappaport 2012; Rappaport et al. 2014). While the failure of GWAS has undermined the importance of the additive contribution of common genetic variants, a deeper focus on the environment-based models has provided empirical evidence of the overwhelming role of the environmental exposures in determining the risk of many common complex diseases.

A number of studies argued that 70 to 90% of the risks of developing several complex diseases like stroke, colon cancer, coronary heart disease, and diabetes, is probably due to environmental factors (Rappaport & Smith 2010; Rappaport 2012). All this empirical evidence pushed scholars to coin in 2005 the new term "exposome" to indicate "life-course environmental exposures (including lifestyle factors) from the prenatal

[5] Heritability is defined as the proportion of total phenotypic variance in a population that is attributable to genetic variation.

[6] "Dominance" refers to interactions between genes at the same locus.

[7] "Epistasis" refers to interactions between genes at different loci.

period onwards" (Wild 2005, p. 1848). The concept of exposome highlights the importance of "a person's lifetime history of all exposures experienced from both external sources (e.g. pollution, radiation, and diet) and internal sources (e.g. inflammation, infection, and the microbiome)" (Lioy & Rappaport 2011, p. 466). While there is a long research tradition on the strong impact on risk of such exposures as diet, lifestyle, and pollution, recently many studies have provided evidence of the role of the *psychosocial* side of the exposome in the determination of the susceptibility to many complex diseases (Wright 2006; Wennerholm et al. 2011).

The exposome can also cause alterations in gene expression by means of epigenetic changes occurring through different patterns of DNA methylation and a wide range of histone modifications. The set of all epigenetic changes is called "epigenome". Epigenetic markings are known to play a causal role in the development of many complex diseases, for instance in carcinogenesis (e.g. Esteller 2007). Epigenetic modifications can be transmitted to offspring in humans in a number of different ways ranging from repetition of similar environment (intrauterine environment included) to direct transmission through the germline (e.g. Jablonka & Raz 2009). Thus epigenetic mechanisms are extremely important in mediating the effects of the early-life exposome on later-life phenotype in the individuals, where the early-life exposome includes both the parental preconceptional exposome and the mother's exposome *in utero* during pregnancy (Hanson et al. 2011).

In short, RNS have so far provided evidence of the primary contribution of the exposome and the epigenome to the risk of complex diseases in the general population, thus supporting the family of environment-based over the family of gene-based models. In the next section we show that self-identified races can be useful proxies for risk-related exposomic and epigenomic variation, since patterns of racial self-identification and patterns of risk-related exposomic and epigenomic variation consistently tend to match each other.

4. *Self-identified races, and risk-related exposomes and epigenomes, as fluid matching patterns*

It is very important to remark that, when scientists employ races, it is by and large *self-identified races* that they are actually making use of. Since differences in racial self-identifications are highly correlated to differences in risk-related exposomes and epigenomes, we do not need to introduce the genetic hypothesis to account for self-identified races as predictors of risks of complex diseases; and we can accommodate for epidemiological data only by stressing the connection, in the contemporary United States, among racial self-identifications on one side, and risk-related exposomes and epigenomes on the other side (see also Gravlee 2009; Kaplan 2010).

Indeed patterns of such heterogeneous risk-related factors as socioeconomic status (Wolfe et al. 2012), access to health care and interactions with health-care providers (Smedley et al. 2009), neighborhood poverty and residential segregation (Cagney et al. 2005), environmental risk exposure (Evans & Kantrowitz 2002), lifestyle (Crespo et al. 2000), diet quality (Hiza et al. 2013), stress levels from perceived discrimination (hence allostatic load and unhealthy epigenetic changes – Seeman et al. 2014), and history of racial discrimination in close ancestors (hence inherited unhealthy epigenetic changes – see e.g. Thayer & Kuzawa 2011 and cites therein) *all shape along racial lines* in the United States.

Self-identified races are what we need to capture this open, complex and highly heterogeneous assortment of environmental, psychological, social, and cultural factors - even more so if we consider how *fluid* self-identified races are. As Saperstein & Penner (2012) remarks, despite the growing consensus about the fluidity of racial identity in the fields of race and ethnicity and social psychology, in practice race is usually treated as a fixed input in most empirical studies of inequality and disparity in the United States. Yet racial identity is more volatile than we can imagine, and tends to vary along with the whole heterogeneous assortment of causal factors contributing to the risks of complex diseases. Change the socioeconomic status, the neighborhood wealth and segregation, or the perceived discrimination of people in the United States, and you will significantly impact their racial self-identification. Thus not only risk-related exposomes and epigenomes shape along racial lines, but also vice versa. If this is true, high is the covariation between self-identified races and risk-related exposomes and epigenomes – so that self-identified races turn out to be (sad as it is) the best proxies for them.

To see how racial self-identification varies along with risk-related exposomes and epigenomes, consider first the case of "multiracial" people, viz. people whose parents self-identify as different races. Their number is constantly increasing all over the world and in particular in the United States, where the population of children with parents of different (self-identified) races has multiplied from 500,000 in 1970 to more than 4 million in 2010 (U.S. Census Bureau 2011). Many studies have shown that in the United States social status

and education influence the choice for a monoracial vs. a multiracial self-identification, in the sense that multiracial individuals of higher status and/or education are more likely than those of lower status and/or education to identify themselves as multiracial (e.g. Townsend et al. 2012). Moreover, Doyle & Kao (2007) found that among those multiracial individuals who identify themselves as belonging to one race, the choice of the race depends on socioeconomic status (besides physical appearance). Interestingly, important exposures for the risk of developing complex diseases – such as socioeconomic status, neighborhood, life quality, stress from discrimination, and the similar – also affect the racial identities assigned to their children by parents identifying as different races. For example, Roth (2005) reported that parents with higher levels of education are more likely to choose a race other than black for their multiracial children; and, if we consider the case of those white-black households that claim a monoracial identity for their children, working-class households are more likely to claim a black identity than middle-class households.

But racial self-identification is influenced by risk-related exposomes and epigenomes *not only* in multiracial individuals. Using data from the 1979 National Longitudinal Survey of Youth (NLSY), Saperstein & Penner (2012) remarks that one in five Americans experienced at least one change in racial classification over a 19-year period; and, the change significantly occurred as a response to changes in social position. So strong is the cultural association between a racial label and specific socioeconomic features, that people often modify their racial self-identification in function of their most recent socioeconomic status. Saperstein & Penner discovered that people having been unemployed for a long period, in poverty, incarcerated, and on welfare are more likely to identify themselves as black, regardless of how they racially self-identified in the previous years. For example, "having ever been unemployed (for more than four months in the same year) more than doubles the odds of identifying as black in 2002, controlling for whether the respondent identified as black in 1979" (Saperstein & Penner 2012, p. 700). Similarly, Doyle & Kao (2007) reports high correlation among changes in racial self-identification and changes in socioeconomic status in "monoracial" individuals.

These studies are very important for understanding why self-identified races reveal as good proxies for the risk of developing complex diseases in the United States. At the population level, fluid patterns of racial self-identification continuously adapt to patterns of heterogeneous assortments of risk-related factors; and vice versa. The result is a consistent tendency for the two configurations to overlap, which endlessly coevolve and reinforce each other. As a consequence, self-identified races turn out to be excellent proxies for complex variation in risk-related environmental, psychological, cultural and social factors that it would be hard to fully account for in any other way.

At the individual level, one could say that racial stereotypes are self-fulfilling prophecies in racist social contexts. A change in racial self-identification may follow a change in the perception of one's socioeconomic status, residential segregation or racial discrimination, but in turn increases the odds of being racially discriminated against[8] and experiencing all health-related negative outcomes. In this sense, racial self-identification can be a factor itself indirectly affecting health. This is a reason why – although one may agree with Root (2005) that whenever differences between members of a population in a biomedical trait are likely to be due to racial discrimination, other-reported rather than self-reported race is prima facie the most appropriate variable to consider – self-reported race turns out to be the best proxy. In fact not only it highly correlates with other-reported race (as said, it is fluid and varies along with the direct material effects of racial discrimination on people's lives; moreover, racial self-identification can have a crucial impact on external social cues (e.g., the way one dresses or behaves) that have primary importance in determining other-reported race; see e.g. Freeman et al. 2011), but it also captures a part of the "racial complexity" that other-reported race does not. To give another example, racial self-identification can be supposed to correlate better than other-reported race with perceived racial discrimination, hence with its biological effects.

5. *The many pathways from racism to health*

We are far from understanding all the causal pathways through which exposure to racial discrimination can affect people's health. Of course some are manifest: for example, racial discrimination contributes to determine socioeconomic status, which in turn affects health in various ways; food quality, exposure to infectious agents and toxic substances, access to health care, housing conditions, education and access to information are obvious mediating factors, as well as material scarcity per se (e.g. Gravlee 2009). But many more causal pathways are much less direct and still poorly understood and investigated.

[8] See, for example, Steele (1997) about how negative cultural images of blacks may adversely affect their academic performance.

For example, stress from discrimination contributes to interparental conflict, children witnessing interparental conflict are placed at heightened risk for emotional problems and for adverse physical and psychological reactions (Rich-Rice 2008), emotion regulation relates to children's early academic success (Graziano et al. 2007), early educational attainment is one important predictor of differing career pathways (Wiesner et al. 2003), and career development impacts all the principal factors being part of the individual risk-related exposome and epigenome (socioeconomic status and racial self-identification included).

Or, consider what can be hidden behind reported evidence that nationalist racial ideology (Sellers & Shelton 2003), racial centrality (Sellers et al. 2003) and ethnic identity (Mossakowski 2003) may be protecting factors buffering the negative impact of discrimination on psychological distress (but see e.g. Cunningham (2012) for opposite evidence), in spite of their being correlated with *increased* perceived discrimination.[9] Sellers & Shelton (2003) suggests that an explanation of this evidence is that nationalist racial ideology (the same holds for racial centrality and ethnic identity) characterizes people for whom experiencing racial hassles may not be unexpected or novel. In other words, these people perceive *more* racial discrimination, but experience *less* psychological distress from it in virtue of their beliefs about the way in which the world works for their racial group being consistent with their personal experiences of discrimination. This latter condition, however, may have various health-related costs. For example, believing that "this is the way in which the world works for blacks" may lower teacher and mother expectations on black youth achievement outcomes, which has a disruptive effect on their actual achievements (Benner & Mistry 2007); as previously stated, worse youth achievement outcomes predict for lower lifetime incomes, which in turn negatively affect risk-related exposome and epigenome in various ways.

Along with similar specific pathways, there is growing evidence that a major causal route exists from ongoing perception of racial discrimination to various health impairments via chronic stress and subsequent high allostatic load, which is defined as "the 'wear and tear' the body experiences when repeated allostatic responses are activated during stressful situations" (Juster et al. 2010; see also Geronimus et al. 2006; Seeman et al. 2014). Many studies found higher values in biomarkers representing primary, secondary or tertiary outcomes in the allostatic load progression to reliably predict incident preterm birth and low birth weight, diabetes, cardiovascular disease, decline in physical functioning, decline in cognitive functioning, and mortality (Juster et al. 2010; Kuzawa & Sweet 2009).

Racial inequalities in allostatic load scores are registered in social contexts such as the contemporary United States that cannot be explained by racial differences in poverty or other relevant features, and are to be thought of as produced by systematic differences in ongoing unhealthy levels of stress due to racial discrimination and racial prejudice (e.g. Geronimus et al. 2006). Therefore allostatic load progression reveals to be a general mechanism through which racism can chronically impact individual health.

The harmful impact of allostatic load can be said partly *transgenerational* insofar as preterm birth and low birth weight – which are among its effects – are associated with high rates of infant death, infant mortality, respiratory and cardiovascular malfunctions, cerebral palsy, intellectual disabilities, vision and hearing complications, and feeding and digestive problems (e.g. Saigal & Doyle 2008). But the unhealthy effects of racism can be transgenerational in a full sense when we consider two further transmission mechanisms: social inheritance of wealth disparities (e.g. Sullivan 2013), and epigenetic inheritance of biological changes. Many studies discovered that numerous pernicious alterations in DNA methylation that result from nutritional stress, psychological stress and environmental toxicants can be transmitted across several generations (Thayer & Kuzawa 2011, and cites therein). In particular, the epigenetic effects of chronic psychological stress caused by racial discrimination and racial prejudice can be passed to and produce health inequalities in subsequent generations (such as, for instance, preterm birth and cardiovascular diseases - see Drake & Seckl 2011; Kuzawa & Sweet 2009), also in absence of continuing environmental stressors.[10]

The abundance of documented causal pathways from racism (even *past* racism) to a large number of health impairments is important under several respects. Firstly, it underscores a significant epistemological dissimilarity between the two main rival explanations of the observed epidemiological differences among self-identified races in the United States, the genetic hypothesis and the "environmental" hypothesis (which explains the epidemiological differences by means of environmental differences stemming from racism and

[9] Centrality, Ideology, and Regard are three dimensions of racial identity measured in 56-item Multidimensional Inventory of Black Identity (MIBI) delineated in the Multidimensional Model of Racial Identity for African Americans (Sellers et al. 1997).

[10] It is worth noting, however, that recent studies on human embryonic stem cells showed that some epigenetic modifications are reversible (Tompkins et al. 2012).

racial prejudices, past and present): while we are rapidly gaining detailed knowledge of the many biological mechanisms involved in the causal pathways connecting racism to health, the genetic variants that have been functionally associated with risks of complex diseases are few and - as previously stated - their relative contribution to the risks is known to be very small with respect to that of the environmental and epigenetic factors (e.g. Vineis et al. 2009). This kind of dissimilarity scores a point for the environmental hypothesis.

Secondly, we understand why we cannot simply eliminate self-identified race as a variable in epidemiology if we want to consider *all* the complex effects of racism on people's health. After all, self-identified race is the only feature to be informative of *all* the complex effects of racism by definition, *included those we do not know*. Since there is a plethora of heterogeneous indirect effects of racism on health, decomposing the "racial effect" and adjusting for *all* the consequences of racial status is highly questionable (e.g. Williams et al. 1997). A fortiori, to adjust self-identified race for socioeconomic factors in order to identify a direct genetic effect that is not mediated by measured covariates appears highly problematic (Kaufman & Cooper 2001). This scores a point against the current use of self-identified races in the genetic hypothesis.

We should not be tempted to hastily conclude that racial discrimination alone can provide a *complete* causal explanation of the differences in the exposome and the epigenome causally responsible for differences in the risk of developing complex diseases among self-identified races in the United States. Agreeing that racism is the main factor shaping the observed biological differences among self-identified races does not amount to agreeing that racism is the only factor. If we imagine a fully racially egalitarian society, we might still presume that a number of race-specific habits would exist (for example, race-specific dietary habits) resulting in risk-related biological differences among self-identified races. However, we must remember that "a fully racially egalitarian society" is a society where not only no racist episode – from the most dramatic to the unnoticeable – is experienced, but also no long-term social, economic or biological (e.g. epigenetic) effect of racism affects people. A fully racism-free society is only one that has been immune to racism permanently or at least for a long time, not just one that recently became so. Sad as it is, this is also a reason for suspecting that so many changes are requested to effectively eradicate racism that perhaps also racial classifications and racial self-identifications would vanish along with it.[11]

It should be added that, while the attention of researchers is normally focused on the negative effects of racism on the health of discriminated racial groups, the symmetrical issue of the positive effects of racism on the health of discriminating racial groups has been systematically neglected. In fact it is arguable that the overall health conditions of privileged and discriminating racial groups are better off than they would be in the absence of racism, just as those of discriminated racial groups are worse off (Jones et al. 2008). Apart from material advantages, health benefits may also stem from perceived ranking at the top of the social hierarchy and consequent psychological advantages. For example, Subramanian et al. (2005) found that absolute health of whites is better where they hold a larger health advantage relative to black counterparts; and Kwate & Goodman (2014) discovered that whites who perceive black families as *less* welcome in their neighborhoods enjoy better health than those who perceive them as more welcome, at least in neighborhoods which are not especially white-segregated and wealthy (the latter features, we may note, signal for major material privileges on their own).

6. *Conclusions*

The role of self-identified races as proxies for a population-specific genetic component to the risk of complex diseases seems to be severely jeopardized by the many difficulties affecting the genetic hypothesis. At the same time, recent biological evidence seems to better support a primary role of non-genetic as opposed to genetic factors in the explanation of the majority of the risks of complex diseases. Nevertheless, there is an alternative promising role for self-identified races within the environment-based models, precisely the role of a proxy for risk-related exposomes and epigenomes.

It is only recently that we began acknowledging the importance and the complexity of the role of the exposome and the epigenome in the explanation of the risks of complex diseases, and systematically exploring them. Many voices were raised stressing that, for this enterprise to be successful, we may need to develop new adequate analytical tools:

---

[11] Or, perhaps more honestly, we should say that we have no idea what "race" would look like (if anything) after the massive and persistent transformations required to fully erase all the epigenetic marks and social disadvantages constituting the legacies of past (and present) racism.

> *There is a desperate need to develop methods with the same precision for an individual's environmental exposure as we have for the individual's genome.* (Wild 2005, p. 1848)

> *Biomedical research is overwhelmingly focused on the gene side of this debate. The tools and knowledge of our nature are far ahead of those for the environment. If we want to focus on the interaction between nature and nurture, we need better ways of cataloguing and integrating the complex exposures and forces that represent nurture.* (Miller & Jones 2014, p. 1)

On the road towards an accurate assessment of the impact of all external and internal exposures, gene-environment interactions, and epigenetic events, self-identified races may be helpful as proxies for population-specific heterogeneous and partially unexplored assortments of risk-related exposures, gene-environment interactions, and epigenetic events which come grouped together, and which we would otherwise be likely to fail to fully consider. Even at a time in which our knowledge of these factors will be much better than today, however, self-identified races may prove indispensable to capture *all* the biological changes due to racism. As long as racism spreads its effects, ranging from the dramatic to the almost unperceivable, self-identified races will be the best proxy for the whole impact of racism upon an individual's health - especially so if we consider that not only risk-related exposomes and epigenomes do shape along racial lines, but also fluid racial self-identifications endlessly follow them and tend to match them, as we have shown. Such continuous coevolution makes self-identified races excellent proxies for risk-related exposomes and epigenomes in the contemporary U.S. We can only add that we do hope self-identified races will cease to have this epistemic power soon - since they can only lose it when racism is permanently erased.

In the present situation, self-identified races may also have a crucial *heuristic* function in epidemiology and biomedical research. They can be used to advance our knowledge of risk-related exposures, gene-environment interactions, and epigenetic events, e.g. by more deeply investigating associations between differences in rates of complex diseases and differences in environmental (external and internal), psychological or social features among individuals who identify themselves as different races. Of course, since the ultimate reason why self-identified races are helpful proxies is the enduring presence of racism in the contemporary U.S. society, the use of self-identified races has a potential important *critical* function as well, since it might remind us that racial discrimination and racial prejudice do have concrete, remarkable biological consequences on people, their minds and their *bodies*. Nonetheless, we should assume neither that this critical function is going to be actually exercised, nor that the total utility of using self-identified races in epidemiology and biomedical research is going to be positive. In fact it might happen that the negative effects overcome the positive effects.

Using self-identified races in biomedical discourse may strengthen the racialization of society and the belief that racial classification is "natural" and inexorable. Since – as previously stressed – self-identified races and inequalities continuously coevolve reinforcing each other and reshaping to match each other, strengthening racial distinctions may result in strengthening racial inequalities – notably racial disparities in health. Another indirect effect might be that of reinforcing unhealthy habits in people perceiving their self-identified race as an immutable burden too heavy and strong to be successfully contrasted. People of disadvantaged racial groups might develop hopelessness and resigned attitudes towards their health conditions and disease risks, and hence negative views of effort and beneficial habits – just as attributing poor performance to a lack of ability which they feel powerless to change rather than to other factors depresses motivation in young students, and believing that intelligence is a fixed trait is a predictor of lower academic success (Blackwell et al. 2007).

All these effects can have disruptive consequences on people's health, enlarging health differences among self-identified races and ironically providing even stronger grounds for using self-identified races in epidemiology and biomedical research (see also Gravlee 2009). There are plenty of such harmful feedback loops that we should identify and carefully consider. We should be prepared to dismiss self-identified races if we suspect that the inequalities caused by racism are reinforced rather than weakened as an overall effect of their employment.

ACKNOWLEDGMENTS

We would like to thank Paolo Francalacci, Guido Barbujani and three anonymous reviewers for helpful comments and suggestions that improved parts of the manuscript. We would also like to thank Roberta Millstein and Rasmus Grønfeldt Winther for their generous support. Part of this research was funded through a Regional grant n. CRP-18397-2009, Regione Sardegna (Italy), and a P.O.R. SARDEGNA F.S.E. 2007–2013 fellowship.

REFERENCES

Aldrich, M. C., Kumar, R., Colangelo, L. A., et al. (2012). Genetic Ancestry-Smoking Interactions and Lung Function in African Americans: A Cohort Study. *PLoS One, 7*, 6, e39541.

Antoniou, A. C., Casadei, S., Heikkinen, T., et al. (2014). Breast-Cancer Risk in Families with Mutations in PALB2. *New England Journal of Medicine, 371*, 6, 497-506.

Baer, R. D., Arteaga, E., Dyer, K., et al. (2013). Concepts of Race and Ethnicity among Health Researchers: Patterns and Implications. *Ethnicity & Health*, *18*, 2, 211-225.

Barbujani, G. (2005). Human Races: Classifying People vs Understanding Diversity. *Current Genomics, 6*, 4, 215-226.

Barbujani, G., Ghirotto, S., & Tassi, F. (2013). Nine Things to Remember about Human Genome Diversity. *Tissue Antigens*, *82*, 3, 155-164.

Benner, A. D., & Mistry, R. S. (2007). Congruence Of Mother and Teacher Educational Expectations and Low-Income Youth's Academic Competence. *Journal of Educational Psychology*, *99*, 1, 140.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, *78*, 1, 246-263.

Bryc, K., Auton, A., Nelson, M. R., et al. (2010). Genome-wide Patterns of Population Structure and Admixture in West Africans and African Americans. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 786-791.

Burchard, E., Ziv, E., Coyle, N., et al. (2003). The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice. *New England Journal of Medicine, 348*, 1170-1175.

Bustamante, C. D., Burchard, E. G., & De La Vega, F. M. (2011). Genomics for the World. *Nature, 475,* 7355, 163-165.

Cagney, K. A., Browning, C. R., & Wen, M. (2005). Racial Disparities in Self-Rated Health at Older Ages: What Difference Does The Neighborhood Make?. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*, 4, S181-S190.

Campbell, M. C., & Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics*, *9*, 403-433.

Chanock, S. J., Manolio, T., Boehnke, M., et al. (2007). Replicating Genotype–Phenotype Associations. *Nature*, *447*, 7145, 655-660.

Crespo, C. J., Smit, E., Andersen, R. E., et al. (2000). Race/Ethnicity, Social Class and Their Relation to Physical Inactivity During Leisure Time. *American Journal of Preventive Medicine*, *18*, 1, 46-53.

Cunningham, J. (2012). Towards a Multidimensional Model of Adaptation for African American Adolescents Exposed to Racial Discrimination. Ph.D. Dissertation Thesis, Loyola University Chicago. Paper 341.

Donovan, B. M. (forthcoming). Putting Humanity back into the Teaching of Human Biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Doyle, J. M., & Kao, G. (2007). Are Racial Identities of Multiracials Stable? Changing Self-Identification among Single and Multiple Race Individuals. *Social Psychology Quarterly*, *70*, 4, 405-423.

Drake, A. J., & Seckl, J. R. (2011). Transmission of Programming Effects Across Generations. *Pediatric Endocrinology Reviews, 9*, 2, 566-578.

Drake, K. A., Galanter, J. M., & Burchard, E. G. (2008). Race, Ethnicity and Social Class and the Complex Etiologies of Asthma. *Pharmacogenomics, 9*, 4, 453-462.

Eeles, R., Goh, C., Castro, E., et al. (2014). The Genetic Epidemiology of Prostate Cancer and its Clinical Implications. *Nature Reviews Urology*, *11*, 1, 18-31.

Eichler, E. E., Flint, J., Gibson, G., et al. (2010). Missing Heritability and Strategies for Finding the Underlying Causes of Complex Disease. *Nature Reviews Genetics*, *11*, 6, 446-450.

Esteller, M. (2007). Cancer Epigenomics: DNA Methylomes and Histone-Modification Maps. *Nature Reviews Genetics*, 8, 4, 286-298.

Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic Status and Health: The Potential Role of Environmental Risk Exposure. *Annual Review of Public Health*, *23*, 1, 303-331.

Fausto-Sterling, A. (2008). The Bare Bones of Race. *Social Studies of Science*, *38*, 5, 657-694.

Feinberg, A. P. (2007). Phenotypic Plasticity and the Epigenetics of Human Disease. *Nature*, *447*, 7143, 433-440.

Feinberg, A. P. (2010). Genome-Scale Approaches to the Epigenetics of Common Human Disease. *Virchows Archiv*, *456*, 1, 13-21.

Fejerman, L., Haiman, C. A., Reich, D., et al. (2009). An Admixture Scan in 1,484 African American Women with Breast Cancer. *Cancer Epidemiology Biomarkers & Prevention, 18*, 11, 3110-3117.

Fejerman, L., Stern, M. C., Ziv, E., et al. (2013). Genetic Ancestry Modifies the Association between Genetic Risk Variants and Breast Cancer Risk among Hispanic and Non-Hispanic White Women. *Carcinogenesis, 34*, 8, 1787-1793.

Frank, R. (2007). What to Make of It? The (Re) Emergence of a Biological Conceptualization of Race in Health Disparities Research. *Social Science & Medicine*, *64*, 10, 1977-1983.

Freeman, J. B., Penner, A. M., Saperstein, et al. (2011). Looking the Part: Social Status Cues Shape Race Perception. *PLoS One*, *6*, 9, e25107.

Fujimura, J. H., Duster, T., & Rajagopalan, R. (2008). Introduction: Race, Genetics, and Disease Questions of Evidence, Matters of Consequence. *Social Studies of Science, 38*, 5, 643-656.

Fujimura, J. H., & Rajagopalan, R. (2011). Different Differences: The Use of 'Genetic Ancestry' versus Race in Biomedical Human Genetic Research. *Social Studies of Science, 41*, 1, 5-30.

Geronimus, A. T., Hicken, M., Keene, D., et al. (2006). "Weathering" and Age Patterns of Allostatic Load Scores among Blacks and Whites in the United States. *American Journal of Public Health*, *96*, 5, 826-833.

Gibson, G. (2012). Rare and Common Variants: Twenty Arguments. *Nature Reviews Genetics, 13*, 2, 135-145.

Gluckman, P. D., Hanson, M. A., & Mitchell, M. D. (2010). Developmental Origins of Health and Disease: Reducing the Burden of Chronic Disease in the Next Generation. *Genome Medicine*, *2*, 14.

Gravlee, C. C. (2009). How Race Becomes Biology: Embodiment of Social Inequality. *American Journal Of Physical Anthropology, 139*, 47–57.

Graziano, P. A., Reavis, R. D., Keane, S. P., et al. (2007). The Role of Emotion Regulation in Children's Early Academic Success. *Journal of School Psychology*, *45,* 1, 3-19.

Hanson, M. A., Low, F. M., & Gluckman, P. D. (2011). Epigenetic Epidemiology: The Rebirth of Soft Inheritance. *Annals of Nutrition and Metabolism, 58*, Suppl. 2, 8-15.

Healthy People 2020. (2014). healthypeople.gov

Hindorff, L. A., Sethupathy, P., Junkins, H. A., et al. (2009). Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits. *Proceedings of the National Academy of Sciences, 106*, 23, 9362-9367.

Hiza, H. A., Casavale, K. O., Guenther, P. M., & Davis, C. A. (2013). Diet Quality of Americans Differs by Age, Sex, Race/Ethnicity, Income, and Education Level. *Journal of the Academy of Nutrition and Dietetics*, *113*, 2, 297-306.

Hung, R. J., McKay, J. D., Gaborieau, V., et al. (2008). A Susceptibility Locus for Lung Cancer Maps to Nicotinic Acetylcholine Receptor Subunit Genes on 15q25. *Nature*, *452*, 7187, 633-637.

Hunt, L. M., & Megyesi, M. S. (2008). The Ambiguous Meanings of the Racial/Ethnic Categories Routinely Used in Human Genetics Research. *Social Science & Medicine*, *66*, 349-361.

Jablonka, E., & Raz, G. (2009). Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution. *The Quarterly Review of Biology*, *84*, 2, 131-176.

Jakobsson, M., Scholz, S. W., Scheet, P., et al. (2008). Genotype, Haplotype and Copy-Number Variation in Worldwide Human Populations. *Nature*, *451*, 7181, 998-1003.

Jones, C. P., Truman, B. I., Elam-Evans, L. D., et al. (2008). Using "Socially Assigned Race" to Probe White Advantages in Health Status. *Ethnicity & Disease, 18*, 496-504.

Juster, R. P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic Load Biomarkers of Chronic Stress and Impact on Health and Cognition. *Neuroscience & Biobehavioral Reviews*, *35*, 1, 2-16.

Kaplan, J. M. (2010). When Socially Determined Categories Make Biological Realities: Understanding Black/White Health Disparities in the U.S.. *The Monist, 93*, 2, 281-297.

Kaplan, J. M., & Winther, R. G. (2013). Prisoners of Abstraction? The Theory and Measure of Genetic Variation, and the Very Concept of "Race". *Biological Theory*, *7*, 4, 401-412.

Kaufman, J. S., & Cooper, R. S. (2001). Commentary: Considerations for Use of Racial/Ethnic Classification in Etiologic Research. *American Journal Of Epidemiology*, *154*, 4, 291-298.

Kistka, Z. A. F., Palomar, L., Lee, K. A., et al. (2007). Racial Disparity in the Frequency of Recurrence of Preterm Birth. *American Journal of Obstetrics and Gynecology*, *196*, 2, 131.e1-131.e6.

Kumar, R., Seibold, M. A., Aldrich, M. C., et al. (2010). Genetic Ancestry and Lung Function Predictions. *New England Journal of Medicine, 363*, 4, 321-330.

Kuzawa, C. W., & Sweet, E. (2009). Epigenetics and the Embodiment of Race: Developmental Origins of US Racial Disparities in Cardiovascular Health. *American Journal of Human Biology*, *21*, 1, 2-15.

Kwate, N. O. A., & Goodman, M. S. (2014). An Empirical Analysis of White Privilege, Social Position and Health. *Social Science & Medicine, 116*, 150-160.

Lee, C. (2009). "Race" and "Ethnicity" in Biomedical Research: How Do Scientists Construct and Explain Differences in Health? *Social Science & Medicine*, *68*, 6, 1183-1190.

Levin, A. M., Wang, Y., Wells, K. E., et al. (2014). Nocturnal Asthma and the Importance of Race/Ethnicity and Genetic Ancestry. *American Journal of Respiratory and Critical Care Medicine*, *190*, 3, 266-273.

Li, J. Z., Absher, D. M., Tang, H., et al. (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science, 319*, 1100-1104.

Lioy, P. J., & Rappaport, S. M. (2011). Exposure Science and the Exposome: An Opportunity for Coherence in the Environmental Health Sciences. *Environmental Health Perspectives, 119*, 11, 466-467.

Manolio T. A., Collins, F. S., Cox, N. J., et al. (2009). Finding the Missing Heritability of Complex Diseases. *Nature, 461*, 7265, 747-753.

Megyesi, M. S., Hunt, L. M., & Brody, H. (2011). A Critical Review of Racial/Ethnic Variables in Osteoporosis and Bone Density Research. *Osteoporosis International*, *22*, 6, 1669-1679.

Miller, G. W. & Jones, D. P. (2014). The Nature of Nurture: Refining the Definition of the Exposome. *Toxicological Sciences, 137,* 1, 1-2.

Mossakowski, K. N. (2003). Coping with Perceived Discrimination: Does Ethnic Identity Protect Mental Health?. *Journal Of Health And Social Behavior, 44*, 318-331.

Paradies, Y., Montoya, M. J., & Fullerton, S. M. (2007). Racialized Genetics and the Study of Complex Diseases: The Thrifty Genotype Revisited. *Perspectives in Biology and Medicine*, *50*, 2, 203-227.

Patel, C. J., Bhattacharya, J., & Butte, A. J. (2010). An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PloS One*, *5*, e10746.

Rappaport, S. M. (2012). Discovering Environmental Causes of Disease. *Journal of Epidemiology and Community Health*, *66*, 2, 99-102.

Rappaport, S. M., & Smith, M. T. (2010). Environment and Disease Risks. *Science*, *330*, 6003, 460-461.

Rappaport, S. M., Barupal, D. K., Wishart, D., et al. (2014). The Blood Exposome and Its Role in Discovering Causes of Disease. *Environmental Health Perspectives, 122*, 8, 769-774.

Rebbeck, T. R., Halbert, C. H., & Sankar, P. (2006). Genetics, Epidemiology, and Cancer Disparities: Is it Black and White? *Journal Of Clinical Oncology*, *24*, 14, 2164-2169.

Rich-Rice, K. L. (2008). *The Effects of Perceived Racial Discrimination on Interparental Conflict and Child Emotional Adjustment in White, Hispanic, and Black Families*. Ph.D. Dissertation Thesis, University of Texas at Arlington.

Root, M. (2005). The Number of Black Widows in the National Academy of Sciences. *Philosophy of Science*, *72*, 5, 1197-1207.

Rosenberg N. A., Pritchard J. K., Weber J. L., et al. (2002). Genetic Structure of Human Populations. *Science*, *298*, 2381-2385.

Roth, W. D. (2005). The End of the One-Drop Rule? Labeling of Multiracial Children in Black Intermarriage. *Sociological Forum 20*, 1, 35-67.

Rotimi, C. N., & Jorde, L. B. (2010). Ancestry and Disease in the Age of Genomic Medicine. *New England Journal of Medicine, 363*, 16, 1551-1558.

Royal, C. D., Novembre, J., Fullerton, S. M., et al. (2010). Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. *American Journal of Human Genetics, 86*, 661–73.

Saigal, S., & Doyle, L. W. (2008). An Overview of Mortality and Sequelae of Preterm Birth from Infancy to Adulthood. *The Lancet*, *371*, 9608, 261-269.

Salas, A., Carracedo, A., Richards, M., et al. (2005). Charting the Ancestry of African Americans. *American Journal of Human Genetics, 77,* 676-680.

Saperstein, A., & Penner, A. M. (2012). Racial Fluidity and Inequality in the United States. *American Journal of Sociology*, *118*, 3, 676-727.

Seeman, M., Merkin, S., Karlamangla, A., et al. (2014). Social Status and Biological Dysregulation: The "Status Syndrome" and Allostatic Load. *Social Science & Medicine, 118*, 143-151.

Sellers, R. M., Rowley, S. J., Chavous, T. M., et al. (1997). Multidimensional Inventory of Black Identity: Preliminary Investigation of Reliability and Construct Validity. *Journal of Personality and Social Psychology, 73*, 805–815.

Sellers, R. M., Caldwell, C. H., Schmeelk-Cone, K. H., et al. (2003). Racial Identity, Racial Discrimination, Perceived Stress, and Psychological Distress among African American Young Adults. *Journal of Health and Social Behavior, 44*, 3, 302-317.

Sellers, R. M., & Shelton, J. N. (2003). The Role of Racial Identity in Perceived Racial Discrimination. *Journal of Personality and Social Psychology*, *84*, 5, 1079-1092.

Spencer, Q. (forthcoming). Philosophy of Race Meets Population Genetics. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Shao, H., Burrage, L. C., Sinasac, D. S., et al. (2008). Genetic Architecture of Complex Traits: Large Phenotypic Effects and Pervasive Epistasis. *Proceedings of the National Academy of Sciences*, *105*, 50, 19910-19914.

Shields, A. E., Fortun, M., Hammonds, E. M., et al. (2005). The Use of Race Variables in Genetic Studies of Complex Traits and the Goal of Reducing Health Disparities: A Transdisciplinary Perspective. *American Psychologist*, *60*, 1, 77-103.

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.). (2009). *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, D.C.: National Academies Press.

Steele, C. M. (1997). A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance. *American Psychologist*, *52*, 6, 613.

Subramanian, S. V., Acevedo-Garcia, D., & Osypuk, T. L. (2005). Racial Residential Segregation and Geographic Heterogeneity in Black/White Disparity in Poor Self-Rated Health in the US: A Multilevel Statistical Analysis. *Social Science & Medicine*, *60*, 8, 1667-1679.

Sullivan, S. (2013). Inheriting Racist Disparities in Health: Epigenetics and the Transgenerational Effects of White Racism. *Critical Philosophy of Race, 1,* 2, 190-218.

Thayer, Z. M., & Kuzawa, C. W. (2011). Biological Memories of Past Environments. *Epigenetics*, *6*, 7, 798-803.

Thomas, D. (2010). Gene–Environment-Wide Association Studies: Emerging Approaches. *Nature Reviews Genetics*, *11*, 4, 259-272.

Tompkins, J. D., Hall, C., Chen, V. C. Y., et al. (2012). Epigenetic Stability, Adaptability, and Reversibility in Human Embryonic Stem Cells. *Proceedings of the National Academy of Sciences*, *109*, 31, 12544-12549.

Townsend, S. S. M., Fryberg, S. A., Wilkins, C. L., & Markus, H. R. (2012). Being Mixed: Who Claims a Biracial Identity? *Cultural Diversity and Ethnic Minority Psychology, 18*, 91–96.

U.S. Census Bureau. (2011). Overview of Race and Hispanic Origin: 2010, *2010 Census Briefs*, report C2010BR-02. Washington, DC: US Government Printing Office.

Vineis, P., Khan, A. E., Vlaanderen, J., et al. (2009). The Impact of New Research Technologies on Our Understanding of Environmental Causes of Disease: the Concept of Clinical Vulnerability. *Environmental Health, 8*, 54.

Weiss, K. M., & Long, J. C. (2009). Non-Darwinian Estimation: My Ancestors, My Genes' Ancestors. *Genome Research*, *19*, 5, 703-710.

Wennerholm, C., Grip, B., Johansson, A., et al. (2011). Cardiovascular Disease Occurrence in Two Close but Different Social Environments. *International Journal of Health Geographics*, *10*, 1.

Wiesner, M., Vondracek, F. W., Capaldi, D. M., et al. (2003). Childhood and Adolescent Predictors of Early Adult Career Pathways. *Journal of Vocational Behavior*, *63*, 3, 305-328.

Wild, C. P. (2005). Complementing the Genome with an "Exposome": The Outstanding Challenge of

Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology Biomarkers & Prevention*, *14*, 8, 1847-1850.

Williams, D. R., Yu, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial Differences in Physical and Mental Health Socio-Economic Status, Stress and Discrimination. *Journal of Health Psychology*, *2*, 3, 335-351.

Wilson, J. F., Weale, M. E., Smith, A. C., et al. (2001). Population Genetic Structure of Variable Drug Response. *Nature Genetics, 29*, 3, 265-269.

Winther, R. G., Giordano, R., Edge, M. D. & Nielsen, R. (forthcoming). The Mind, the Lab, and the Field: Three Kinds of Populations in Scientific Practice. *Studies in History and Philosophy of Biological and Biomedical Sciences*.

Wolfe, B., Evans, W., & Seeman, T.E. (2012). *The Biological Consequences of Socioeconomic Inequalities*. New York: Russell Sage Foundation.

Wray, N. R., Purcell, S. M. & Visscher, P. M. (2011). Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results. *PLoS Biology, 9*, e1000579.

Wright, R. J. (2006). Health Effects of Socially Toxic Neighborhoods: the Violence and Urban Asthma Paradigm. *Clinics in Chest Medicine, 27*, 3, 413-421.