Temporal changes in genetic admixture are linked to heterozygosity and health diagnoses in humans

Brian S. Mautz,[1,2] Jacklyn N. Hellwege,[1,2] Chun Li[4], Yaomin Xu,[5] Siwei Zhang,[5] Joshua C. Denny,[2,7] Dan M. Roden,[1,2,6,7] Tracy L. McGregor,[8,9] Digna R. Velez Edwards,[1,3,7,9] and Todd L. Edwards[1,2,9*]

[1] Vanderbilt Epidemiology Center and Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN, USA

[2] Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA

[3] Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University School of Medicine, Nashville, TN, USA

[4] Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA

[5] Department of Biostatistics, Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, TN, USA

[6] Department of Pharmacology and Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA

[7] Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA

[8] Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN, USA

[9]Senior author

[*]Correspondence: todd.l.edwards@vanderbilt.edu

Abstract

Reproduction between individuals from different ancestral populations creates genetically admixed offspring. Admixture can have positive and negative impacts on individual health, feeding back to population health. Historical and forced migrations, and recent mobility, have brought formerly disparate populations of humans together. Here we sought to better understand how temporal changes in genetic admixture influence levels of heterozygosity and health outcomes. We evaluated variation in ancestry over 100 birth years in 35,842 individuals from a genetic database linked to health records in a population in the Southeastern United States. Analysis of 2,678 ancestrally informative markers revealed increased admixture and heterozygosity for all clinically-defined race groups since 1990. Most groups also exhibited increasing long-range linkage disequilibrium over time. A phenome-wide association study of clinical outcomes detected protective associations with female reproductive disorders and increased risk for diseases with links to autoimmunity dysfunction. These mixed effects have important ramifications for human health.

Keywords: electronic health records, genetic admixture, heterozygosity fitness correlation, health outcomes, long-range linkage disequilibrium

Introduction

Understanding temporal changes in heterozygosity has important implications for individual- and population-level health in humans that remain unexplored. The heterozygosity-fitness correlation (HFC) is a measure used to quantify the extent to which proxies of fitness are positively correlated with the level of heterozygosity across sets of genetic loci or the whole genome (David, 1998; Head et al., 2017; Szulkin et al., 2010). Evidence linking fitness components to levels of heterozygosity has been found in animals (Keller and Waller, 2002), but results are equivocal and controversial (Chapman et al., 2009; Szulkin et al., 2010). HFCs arise through some level of inbreeding depression or can be generated through immigration and population admixture (reviewed in Szulkin et al., 2010). Both positive and negative effects on individual and population health can occur.

Three factors make human population genetic data ideal to study the connection between heterozygosity and health. First, the human genome is sequenced and genetic databases with large numbers of genetic loci and individuals are readily available. Second, there are substantial numbers of subjects with genetic data linked to electronic health records (EHR) (Roden et al., 2008; Sudlow et al., 2015). Third, the diseases and their etiologies used in health records are known in much more detail and well categorized than most other species, facilitating the estimation the relationship between heterozygosity and changes in (disease) phenotypes used as fitness proxies.

Additionally, we have detailed information on the movements of many human populations. Geographically disparate human populations have come into contact through

65 migration, as well as more recent mobility and urbanization, increasing the opportunity for
66 genetic admixture and HFCs to occur. Genetic admixture has previously been used to identify
67 migration patterns across several humans populations (Baharian et al., 2016; Bryc et al., 2010,
68 2015; Han et al., 2017; Moreno-Estrada et al., 2014; Tishkoff et al., 2009; Wang et al., 2013) and
69 the genetic basis of diseases (Franceschini et al., 2013; Kato et al., 2015; Reich and Patterson,
70 2005). Two studies have shown temporal increases in heterozygosity due to urbanization, one
71 in a Croatian population (Rudan et al., 2008) and one in a US population of European ancestry
72 (Nalls et al., 2009). However, these studies of admixture have not connected migratory or
73 urbanization patterns to health outcomes.

74 Finally, accruing evidence in humans suggests HFCs occur. For example, several studies
75 have linked levels of heterozygosity in non-admixed populations to specific disease states
76 (Bittles and Black, 2010; Campbell et al., 2007, 2009; Pemberton and Szpiech, 2018).
77 Additionally, a meta-analysis of populations of European American and African American
78 ancestry found a positive association between levels of heterozygosity and mortality in humans
79 (Bihlmeyer et al., 2014). However, these studies are limited in scope, investigating single
80 outcomes and ignoring the impact of temporal trends in admixture and heterozygosity on
81 epidemiological outcomes.

82 Here, we link temporal changes in genetic admixture to patterns of heterozygosity and
83 epidemiological outcomes across a broad array of human diseases. We used BioVU, a
84 biorepository of genetic data linked to EHR (Roden et al., 2008) to investigate temporal changes
85 in genetic admixture and heterozygosity in a population in the Southeastern United States.
86 Though race has little biological meaning, we used race as instrumental groupings (Maglo et al.,

87 2016) to reflect corresponding socio-cultural dynamics in this geographic area and investigate
88 differential changes to admixture and heterozygosity within these groupings. In addition, we
89 used a phenome-wide association study (PheWAS, Denny et al., 2010), to connect genetic data
90 with the clinical phenome representing an entire collection of disease outcomes associated
91 with EHRs in BioVU.

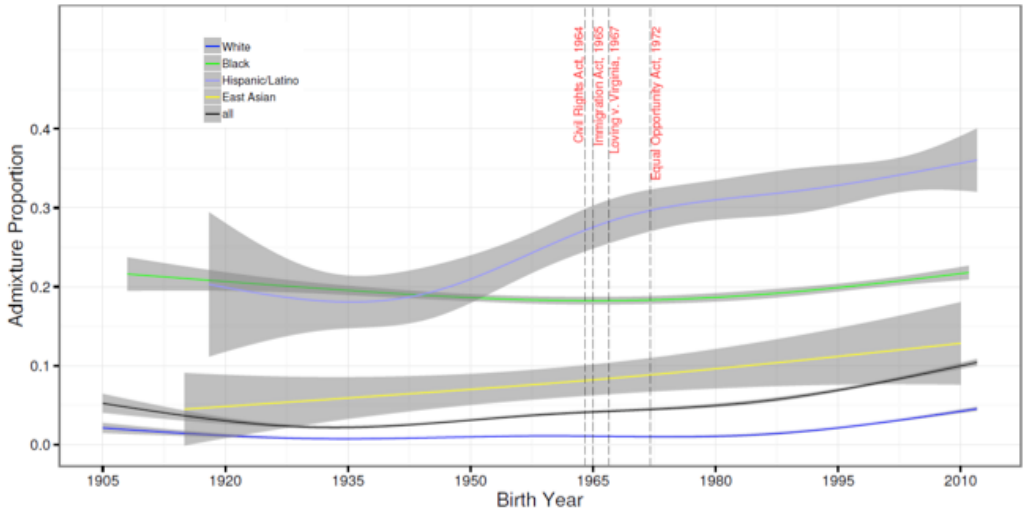
92

93 **Results**

94 We identified 2,678 ancestry-informative markers (AIMs) from genetic data in BioVU.
95 These AIMs were from ExomeChip data in a cohort of 28,723 Whites, 4,129 Blacks, 550
96 Hispanic/Latinos, and 270 Asians, based on the EHR third-party race designation (see methods).
97 Sample population demographics by EHR race and birth decade are in presented in
98 Supplementary Table S1-2 and Figure S1. After combining our data with the 1000 Genomes as a
99 reference group (Durbin et al., 2010), we calculated principal components to identify patterns
100 of ancestry in each individual. We used the ancestral classifications to test for temporal trends
101 in mean and variance in admixture proportion.

102 Analysis of temporal trends in genetic admixture showed an increase in ancestral
103 diversity with time (Figure 1; see Table S2). The admixture proportion consistently increased
104 with younger ages in Asian and Hispanic/Latino groups. In Whites, the level of admixture
105 remained stable until approximately 1990 when it began to increase. Blacks had a decrease in
106 admixture proportion in those born the early- to mid-20th century, but an increase after the
107 1980's. In addition to the distinct clusters in the principal component plots representing

108 individuals who are identified as either White and Black and who contain both European and
109 African ancestries, the overall proportions of ancestry across all recorded race groups exhibit
110 increasing admixture in younger individuals (Figure S3-S7). Additionally, a significant increase in
111 the variance of admixture proportions over time was observed (variance coefficient = 0.0193 ±
112 0.0009 [SE], p-value < 1.44×10^{-100}), indicating that there is a linear increase in variance of the
113 admixture proportion of 0.0193 with every birth year. This is consistent with recent and
114 ongoing admixture (Verdu and Rosenberg, 2011).



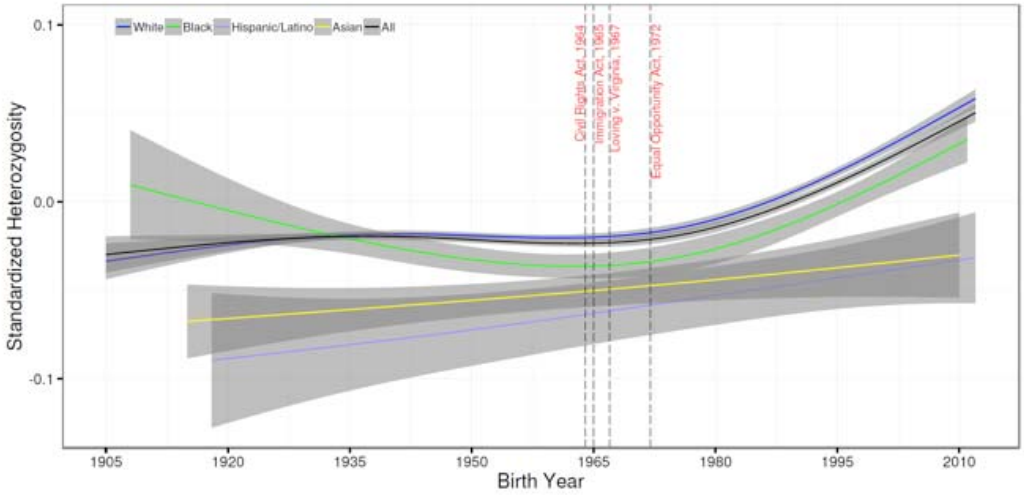
115

116 **Figure 1**. Admixture proportion plotted against birth year. Admixture proportion is defined as (1
117 - maximum ancestry proportion). The smoothing curves are obtained using the generalized
118 additive model method with a cubic spline basis implemented in R package mgcv and plotted
119 using R package ggplot2. Cohorts are grouped by race and 95% confidence intervals are
120 included.

121

122 We detect similar patterns in three additional sources. First, we calculated heterozygosity
123 across AIMs and tested for temporal trends to compare with patterns of genetic admixture.
124 There was strong support for increasing heterozygosity in younger individuals (Figure 2; Table

125 S3 and Figure S8). The timing of inflection for increased standardized heterozygosity varied
126 between the race groups, but the data indicated that ancestry diversity has accelerated rapidly
127 in Asian, Black, and White cohorts since about 1980, while Hispanics/Latinos have exhibited a
128 relatively steady rate of increasing diversity since the 1940s. This reflects the increasing number
129 of children born to couples of predominantly different ancestral backgrounds over the past few
130 decades.



131

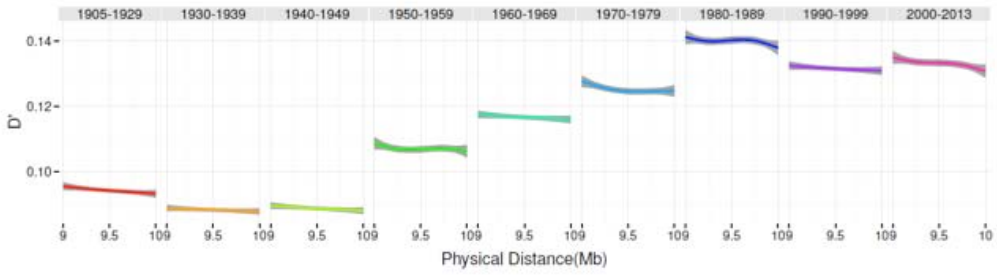
132 **Figure 2**. Standardized heterozygosity plotted against birth year. The y-axis depicts the
133 standardized heterozygosity of individuals with birth year on the x-axis. Cohorts are grouped by
134 stated race and 95% confidence intervals are indicated.

135

136 Second, we verified patterns by estimating pairwise long-range linkage disequilibrium in our
137 dataset. Using common single nucleotide polymorphisms (SNPs) (minor allele frequency [MAF]
138 > 5%) from the genotyping array, we calculated pairwise D' using Haploview (Barrett et al.,
139 2005) and plotted D' against physical distance for all pairs of SNPs between 9–10 megabases of
140 each other (Figure 3; Table S4). Results show a small drop in long range disequilibrium (LRLD) in

141 individuals born before the 1950s, followed by significant steady increases in the 1950s through

142 the 1980s and subtle fluctuation at higher levels in the 1990s to 2010s.



143

144 **Figure 3**. Pairwise D' for SNPs between 9-10 Mb for all common SNPs on the exome array in all
145 study individuals stratified by intervals of birth decade. Pairwise D' for all common SNPs
146 between 9 - 10 megabases apart for individuals are presented on the y-axis and physical
147 distance in Mb is presented on the x-axis. The plots are paneled by decades of birth to depict
148 relative changes over time.

149

150 Third, we estimated changes in the number of ethnicities in self-reported ancestry in the

151 2013 American Communities Survey. Respondents were instructed to select

152 Hispanic/Latino/Spanish status and all applicable races for each individual in the household

153 (Ruggles, 2014). We analyzed the average number of race categories selected for each

154 individual by birth year and stratified these results within race categories for Tennessee (Figure

155 S9), the East South-Central Region which includes Tennessee, Alabama, Kentucky, and

156 Mississippi (Figure S10), and the entire United States (Figure S11). The results from all regions

157 are concordant, indicating that younger individuals are more likely to indicate multiple races.

158 The inflection point appears to have been earlier in the Asian race category but is demonstrated

159 in all race groups by the mid-1960s in the entire US sample.

Changes in health diagnoses with admixture

We evaluated the association between individual heterozygosity and disease diagnoses using a phenome-wide association study approach (PheWAS, Denny et al., 2010). Increasing genetic admixture resulted in fewer diagnoses of female reproductive traits across all data (Table 1). These results remain statistically significant when considering adults only (Table S5) and after correction for multiple tests. Phenotype codes for "disorders of menstruation and other abnormal bleeding from female genital tract" and "irregular menstrual cycle/bleeding" were significantly associated with protection by increasing heterozygosity (p-value = 7.21×10^{-6} and 4.37×10^{-5}, respectively; Table 1, Figure S12). Other protective findings were also gynecological in nature, including cervical cancer/dysplasia and abnormal Papanicolaou smear results. Significant phenotypes in adults were predominantly detected for females. Outside genitourinary findings, other nominally significant associations (Bonferroni significant $< p \leq 0.05$) show increased risk with genetic admixture and include atopic dermatitis, AV Block, obstructive asthma, Sicca syndrome.

Table 1. Results from the phenome-wide association study of heterozygosity and clinical outcomes for full sample.

PheCode	Phenotype	P-value	OR (95% Confidence Interval)
626	Disorders of menstruation and other abnormal bleeding from female genital tract	7.21×10^{-6}	0.37 (0.24 – 0.57)
626.1	Irregular menstrual cycle/bleeding	4.37×10^{-5}	0.37 (0.23 – 0.60)
939	Atopic/contact dermatitis due to other or unspecified	2.86×10^{-4}	1.82 (1.32 – 2.52)
180	Cervical cancer and dysplasia	4.02×10^{-4}	0.19 (0.08 – 0.48)
792.1	Papanicolaou smear of cervix or vagina with atypical squamous cells	5.36×10^{-4}	0.21 (0.09 – 0.51)

180.3	Cervical intraepithelial neoplasia [CIN] [Cervical dysplasia]	6.24×10^{-4}	0.15 (0.05 – 0.45)
426.2	Atrioventricular [AV] block	7.62×10^{-4}	2.97 (1.58 – 5.61)
495.11	Chronic obstructive asthma with exacerbation	8.13×10^{-4}	4.63 (1.89 – 11.34)
785	Abdominal pain	1.04×10^{-3}	0.70 (0.57 – 0.87)
792	Abnormal Papanicolaou smear of cervix and cervical Human Papilloma Virus (HPV)	1.04×10^{-3}	0.33 (0.17 – 0.64)
709.2	Sicca syndrome	1.66×10^{-3}	4.38 (1.75 – 10.99)
318	Tobacco use disorder	1.75×10^{-3}	0.48 (0.30 – 0.76)
626.12	Excessive or frequent menstruation	1.80×10^{-3}	0.28 (0.12 – 0.62)
628	Ovarian cyst	2.83×10^{-3}	0.23 (0.09 – 0.60)
426.24	Atrioventricular block, complete	3.57×10^{-3}	3.30 (1.48 – 7.38)

Discussion

We found increasing genetic admixture and heterozygosity through time in a population of the Southeastern United States. Comparing these changes to socio-cultural shifts in this geographic area puts these results in context. In the Southern United States from the Civil War until the mid-1960's, laws and policies enforced segregation of populations of European and African ancestry. Consistent with these socio-cultural boundaries, there is little change in admixture through the 1960's. Additionally, despite legal rulings and perceived socio-cultural transformation, there remained a very slow increase in admixture and heterozygosity for an additional 20-30 years. For individuals with an EHR designation of White, the mean proportion of European ancestry was greater than 98% until the 1990's but decreased to 92% in the youngest cohort as the proportion of African ancestry increased to 6% (Figure S4). In individuals designated as Black in the EHR, the mean African proportion was approximately 80% in those born before the 1990's but decreased to 77% in those born more recently (Figure S5). The largest change of any group was in those identified as Hispanic/Latino. The European ancestry

192 proportion in these individuals decreased by 15%, from approximately 75% in those born
193 before the 1950's to under 60% after the 1980's. This decrease in European ancestry was
194 associated with an increase in the East Asian ancestry proportion, likely a proxy for Native
195 American ancestry, while the African proportion remained under 10% (Figure S6). The
196 continental ancestry proportions in EHR-designated Asians were more difficult to discern
197 because of the small cohort size, although we did observe a slight decrease in the East Asian
198 proportion over time (Figure S7). These trends are consistent with previous reported trends in
199 decreases in homozygosity in other populations (Nalls et al., 2009; Rudan et al., 2008).

200 Long-range linkage disequilibrium occurs when subdivided populations mix (Koch et al.,
201 2013; Nei and Li, 1973). Consistent with this and with patterns we detected in admixture and
202 heterozygosity, we find that LRLD changes through time (Figure 3). There was small drop in
203 LRLD for older individuals born before 1950. This was followed by a consistent increase in LRLD
204 in individuals born up to 1980. Subtle fluctuations at higher LRDL levels occurs in those
205 individuals born between 1990 and 2010. It should be noted, however, that we cannot
206 necessarily rule out other possible sources of LRLD (e.g. drift, epistatic selection, hitchhiking as
207 in Koch et al., 2013), though they seem unlikely given the recent nature of admixture and
208 formation of LRLD.

209 Our results are qualitatively echoed in the 2013 American Communities Survey. Younger
210 individuals, across multiple geographic areas (Tennessee, East South-Central Region, United
211 States) are more likely to indicate multiple races when asked about their ancestral origins
212 (Figure S9-11). Across race group the inflection point occurs by the mid-1960's of the entire US
213 sample, though appears earlier for the Asian race category.

214

215 *Changes in health diagnoses with admixture*

216 We show that changes in population genetic parameters have important consequences
217 for population health (Table 1; Table S5 and Figure S12), which have not been previously
218 documented. There were several statistically significant associations of genetic admixture with
219 adult female genitourinary diagnosis codes (e.g. irregular menstrual cycle/bleeding; cervical
220 cancer/dysplasia). These findings linking admixture to protection from menstruation and
221 gynecological abnormalities suggests that ancestral diversity is related to risk of disorders that
222 could affect reproduction. This is consistent with long-standing models that propose ancestral
223 diversity and heterozygosity is related to fertility in plants and animals, as well as other
224 beneficial characteristics in offspring (Dickerson, 1973; Keller and Waller, 2002; Lamkey and
225 Edwards, 1999; Shull, 1948; Szulkin et al., 2010).

226 The changes in reproductive diagnoses were detected predominantly for females,
227 suggesting a largely sex-specific response to changes in admixture and heterozygosity. In some
228 animals, there are sex-specific responses in reproduction to inbreeding (Keller and Waller,
229 2002), providing some circumstantial support for our findings. However, the sex-specific
230 response we detected could also be a results of gender differences in other factors. For
231 example, men are less likely to seek medical attention and report medical issues (Galdas et al.,
232 2005; Mansfield et al., 2003). Alternatively, male reproductive traits, such as sperm quality,
233 might not be routinely checked and reported as with female reproductive parameters. More

234 research is needed to disentangle biological and sociological factors driving the sex-specific
235 responses to determine the source of these trends.

236 Other interesting patterns emerge when considering all data and nominally significant
237 PheWAS results (i.e. $p < 0.05$). Several of these disease phenotypes, show positive associations
238 with increasing genetic admixture. Importantly, each of these diseases have at least suggestive
239 links to autoimmunity issues (*atopic dermatitis*: Mittermann et al., 2004, *AV block*: Buyon et al.,
240 1998; Villuendas et al., 2014, *asthma*: Barnes, 2008; Tedeschi and Asero, 2008, *Sicca/Sjögren*
241 *syndrome*: Brito-Zerón et al., 2016). The nominally significant diseases detected here also have
242 reported connections with each other in the literature. For example, atopic dermatitis and
243 asthma have a similar mechanistic basis via allergen exposure through the epidermis (Spergel,
244 2010). Similarly, autoimmune congenital heart block is linked to Ro and La autoantigens,
245 Sjögren-syndrome-related antigen A and B respectively, which cross the placenta from the
246 mother to the fetus to cause cardiac abnormalities and cutaneous rashes (Brito-Zerón et al.,
247 2015). These patterns suggest increasing heterozygosity are linked to increases in activity in the
248 immune system. Our results show continued increases in genetic admixture with time. If
249 heterozygosity is positively linked to autoimmunity-related disease, then we would predict
250 continued increases in prevalence of these types of disease with time. Future research should
251 address these intriguing immunity-disease relationships to determine the validity and
252 consistency of these patterns within and across human populations.

253 These results have important ramifications in several areas. First, this work has
254 implications for genetic association study designs, disease prevalence and disparities, and
255 classification of individuals for their medical care. Frequently genetic analyses are stratified by

256 ancestry or race, and our data show that simple categorical race classifiers, whether by self-
257 report or derived from other data, are becoming less accurate over time. When conducting
258 genetic studies in children, researchers may need to collect the race(s) of the parents or even
259 grandparents to allow for appropriate classification, or consider AIMs to estimate ancestry
260 proportions (Kodaman et al., 2013; Ruiz-Narvá Ez et al., 2011). This increase in ancestral
261 diversity of the younger population may also enable admixture mapping studies in populations
262 not traditionally considered for these types of studies.

263 Second, understanding race-associated factors in patients with complex ancestries may
264 be increasingly important for effectively delivering precision medical care, which is a major
265 National Institutes of Health (NIH) goal and one of the key promises of the genetics research
266 enterprise in the United States. Genetic information has been used to investigate the roots of
267 human evolution and subsequent diversity. These same approaches can be applied to analyze
268 ongoing trends in regional and national demographics which have significant implications for
269 scientific, societal, and governmental policies.

270 Finally, these results have important implications for HFCs and their use in biological
271 research (Keller and Waller, 2002; Chapman et al., 2009; Szulkin et al. 2010). We detected
272 positive changes in reproductive traits in females, providing evidence that genetic admixture
273 and heterozygosity are linked to potential fitness-related components. Though there were
274 protective effects for reproductive diagnoses, our results provide additional details showing
275 increased risk associated with other disease states that might be linked with the immune
276 system as genetic admixture increases and that HFCs might be sex-specific. The effect of

277 changes to levels on health and fitness related parameters are complex, so HFCs should be
278 interpreted with caution and more detail is needed for proper interpretation.
279

280 **Materials and Methods**

281 Individuals were selected from the BioVU DNA Repository which links clinical data from
282 de-identified electronic medical records to DNA specimens obtained from patients at Vanderbilt
283 University Medical Center (VUMC) (Roden et al., 2008). Each individual is assigned a single race
284 in the Electronic Heath Record (EHR) of White, Black, Asian, Pacific Islander, American
285 Indian/Alaska Native, or declined/unknown, and an ethnicity of Hispanic/Latino, Not
286 Hispanic/Latino, or declined/unknown. BioVU also contains third-party-designated race, which
287 is a good predictor of self-reported ethnicity/ancestry in this database (Dumitrescu et al., 2010).
288 This study of de-identified data was determined to be non-human subject research by the
289 institutional review board (IRB) of Vanderbilt University, Nashville, TN.
290

291 **DNA Extraction and Genotyping**

292 All DNA samples were isolated from whole blood using the Autopure LS system (QIAGEN
293 Inc., Valencia, CA). Genomic DNA was quantitated via an ND-8000 spectrophotometer and DNA
294 quality was evaluated via gel electrophoresis. Individuals were genotyped using the Illumina
295 Infinium HumanExome Array [12v1-1] (Illumina Inc., San Diego, CA). The data were processed
296 for genotype calling using Illumina's Genome Studio (Illumina Inc., San Diego, CA).
297

298 **Genotyping Quality Control**

299 Data on 240,117 SNPs and 35,842 individuals (16,289 males, 19,552 females) were
300 available prior to implementation of quality control (QC) measures. No individuals were
301 excluded for low genotyping efficiency (<98%). 6,599 SNPs were excluded for low genotyping
302 efficiency (<98%) and 71,667 SNPs were monomorphic. Twenty-six individuals (14 EHR males,
303 12 EHR females) were excluded for inconsistent genetic and database sex. After QC, 163,135
304 SNPs remained for analyses in 35,456 individuals. No SNPs were removed for deviations from
305 Hardy-Weinberg equilibrium.
306
307 **Quantification and Statistical Analyses**
308 Descriptive statistics of demographic and clinical characteristics were expressed as
309 mean with standard deviation or median with interquartile range for continuous covariates and
310 as frequencies or proportions for categorical data using SPSS statistical software (IBM
311 Corporation, Armonk, NY).
312 A subset of 2,678 ancestry-informative markers (AIMs) were selected for subsequent
313 analysis. We selected AIMs from the ExomeChip content, which were selected to have strong
314 differences between African and European ancestry populations as well as Native American and
315 European ancestry populations. Average differences in allele frequency, Δ_p, between 1000
316 Genomes (Price et al., 2006) continental populations (European, Asian, Americas, African) for
317 the AIMs are presented in Table S6. AIMs were used rather than using pruned SNP data due to
318 the genotype data coming from an ExomeChip platform, which was designed with a panel of
319 AIMs to enable evaluation of ancestry.

320 EIGENSTRAT v6.0.1 software was used to conduct principal components analysis (PCA)
321 to estimate continuous axes of ancestry from AIMs running all populations together (Price et
322 al., 2006).
323 STRUCTURE software v2.3.3 (Falush et al.; Pritchard et al., 2000) was used to quantify
324 ancestry in combined study and 1000 Genomes Project (Price et al., 2006) individuals using the
325 AIMs. We estimated proportions of ancestry assuming ancestral clusters (*K*) ranging from one
326 to 16, where 16 is the number of sub-populations in the 1000 Genomes data plus two. We
327 assumed unlinked SNPs and used 5,000 iterations of burn-in and 10,000 iterations for analysis
328 without providing population information to the software. We observed that the –log-
329 likelihood of the data given *K* did not vary significantly for *K*'s greater than three and observed
330 that *K*'s greater than three primarily subdivided the European populations (data not shown).
331 The three STRUCTURE clusters corresponded to African, European, and Asian ancestry based on
332 comparisons to the 1000 Genomes reference data.
333 Each of the three derived proportions of continental ancestry from STRUCTURE were
334 regressed onto birth year using generalized additive models with integrated smoothness
335 estimation (GAM) (Wood, 2010) implemented in the R package mgcv in all study individuals,
336 Non-Hispanic Whites , Non-Hispanic Blacks, Hispanics/Latinos, and Non-Hispanic Asians. We
337 derived admixture proportion, α, for the *i*th individual using the formula α_i = 1 – maximum
338 (%European, %African, %East Asian). We regressed α_i onto birth year using generalized additive
339 models with integrated smoothness estimation for all study individuals, Non-Hispanic Whites,
340 Non-Hispanic Blacks, Non-Hispanic Asians, and Hispanics or Latinos. Proportions of continental

ancestry in the Non-Hispanic Whites, Non-Hispanic Blacks, Non-Hispanic Asians, and Hispanics or Latinos by birth decade are presented.

It has been previously shown by Verdu and Rosenberg (2011) that when parental populations stop contributing to admixture, that the variance of admixture proportions decreases rapidly, and when they continue to contribute to the admixture, that the variance of admixture proportions increases over time. To test null hypothesis that the observed levels of admixture in our data were not due to ongoing and increasing rates of admixture, we analyzed the association between the variance of α and birth year. We used the software package MVtest and modeled the admixture proportion variance as a log-linear function of birth year and five principal components of ancestry using estimating equations.

Analysis of admixture proportion variance over time

We model admixture proportion mean and variance simultaneously as functions of birth year and covariates. Specifically, let m_i be the mean and σ_i^2 be the variance of the trait Y_i for the *i*-th individual. We model them as

$$m_i = \beta_0 + \beta_g G_i + \sum_{j=1}^{p} \beta_j X_{ij},$$

and

$$\ln(\sigma_i^2) = \gamma_0 + \gamma_g G_i + \sum_{j=1}^{p} \gamma_j X_{ij} \ .$$

where G_i is the birth year or variable of interest for the *i*-th individual, and X_{i1}, ..., X_{ip} are *p* covariates. In this model the variance is monotonic with respect to birth year, an assumption that holds in most circumstances. The parameters are estimated simultaneously. This framework allows for testing of the null hypothesis of no effect on mean, variance, or both for

any term in the model. These correspond to a mean test with null H_0: $\beta_g = 0$, a variance test with H_0: $\gamma_g = 0$, both having one degree of freedom (DF), and a 2-DF test with H_0: $\beta_g = 0, \gamma_g = 0$.

Model fitting with estimating equations

The parameters are estimated through the estimating equations approach, which does not require a full specification of the outcome distribution, but only a few constraints for the parameters of interest. These constraints are often written as equations, and the parameter estimates can be obtained by solving the equations. The asymptotic distribution for the parameter estimates can be derived (Stefanski and Boos, 2002). Specifically, suppose the random variable Y_i has mean $m_i = \beta_0 + \beta_g G_i + \sum_{j=1}^{p} \beta_j X_{ij}$, and log-variance $\ln(\sigma_i^2) = \gamma_0 + \gamma_g G_i + \sum_{j=1}^{p} \gamma_j X_{ij}$. There are $k = 2(p+2)$ parameters, which can be written as a vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$, where $\boldsymbol{\beta} = (\beta_0, \beta_g, \beta_1, \dots, \beta_p)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_g, \gamma_1, \dots, \gamma_p)$. Let y_i and $\boldsymbol{x}_i = (1, g_i, x_{i1}, \dots, x_{ip})^T$ be the observed values for subject *i*. If we had assumed normality for the outcome, the log-likelihood for observation *i* would have been

$$l_i(\boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{x}_i - \frac{(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)^2}{2\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)},$$

for which the partial derivatives with respect to the parameters $\boldsymbol{\theta}$ is a *k*-vector,

$$\Psi_i(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial l_i}{\partial \boldsymbol{\beta}} \\ \frac{\partial l_i}{\partial \boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \frac{y_i - \boldsymbol{\beta}'\boldsymbol{x}_i}{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)}\boldsymbol{x}_i \\ \frac{1}{2}\left[\frac{(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)^2}{\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i)} - 1\right]\boldsymbol{x}_i \end{pmatrix},$$

and the maximum likelihood estimates of the parameters could have been obtained by solving the *k* equations $\sum_{i=1}^{n} \Psi_i(\boldsymbol{\theta}) = 0$. This motivated us to use these *k* equations,

$$\sum_{i=1}^{n} \Psi_i(\boldsymbol{\theta}) = 0,$$

as the starting point for our estimating equations approach to obtain parameter estimates $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}})$. If normality holds, then $\widehat{\boldsymbol{\theta}}$ are the maximum likelihood estimates. Note that although the estimating equations were motivated by the Gaussian likelihood, one can always start from these equations to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$, whether normality holds or not, and proceed with statistical inference using the M-estimation theory (Stefanski and Boos, 2002). This is a major advantage for using estimating equations. The partial derivative of $\Psi_i(\boldsymbol{\theta})$ is a $k \times k$ matrix, denoted as $\psi_i(\boldsymbol{\theta})$. Using the M-estimation theory, we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N(0, V),$$

where the $k \times k$ covariance matrix V can be estimated as $A^{-1}B(A^{-1})^T$ with $A = -\frac{1}{n}\sum_{i=1}^{n} \psi_i(\widehat{\boldsymbol{\theta}})$ and $B = \frac{1}{n}\sum_{i=1}^{n} \Psi_i(\widehat{\boldsymbol{\theta}})\Psi_i(\widehat{\boldsymbol{\theta}})^T$.

If our interest is on the effect of G, the asymptotic result for the joint distribution for the parameter estimates $\hat{\beta}_g$ and $\hat{\gamma}_g$ is

$$\sqrt{n}\left[\begin{pmatrix}\hat{\beta}_g \\ \hat{\gamma}_g\end{pmatrix} - \begin{pmatrix}\beta_g \\ \gamma_g\end{pmatrix}\right] \xrightarrow{d} N(0, V_2),$$

where V_2 is the corresponding 2×2 submatrix of V, with diagonal values denoted as $\hat{\sigma}^2_{\beta_g}$ and $\hat{\sigma}^2_{\gamma_g}$, respectively. A mean test (H_0: $\beta_g = 0$) can be performed by comparing $\sqrt{n}\hat{\beta}_g$ with $N\left(0, \hat{\sigma}^2_{\beta_g}\right)$, and similarly, a variance test (H_0: $\gamma_g = 0$) by comparing $\sqrt{n}\hat{\gamma}_g$ with $N\left(0, \hat{\sigma}^2_{\gamma_g}\right)$. A 2-DF joint test ($H_0$: $\beta_g = 0, \gamma_g = 0$) can be performed by comparing $n(\hat{\beta}_g, \hat{\gamma}_g)V_2^{-1}\begin{pmatrix}\hat{\beta}_g \\ \hat{\gamma}_g\end{pmatrix}$ with a chi-squared distribution with two degrees of freedom. MVtest software for genetic analysis of SNP

397 data or general analysis of variables is freely available at https://github.com/edwards-
398 lab/MVtest.
399
400 **Heterozygosity Analysis**
401 Standardized measures of heterozygosity among the AIMs were calculated in order to
402 evaluate trends in heterozygosity over time relative to expectations. We first estimated the
403 expected number of heterozygous genotypes in an individual in the *k*-th subpopulation as
404 (1) $H_k = \sum_i 2p_{ik}q_{ik}$,
405 where the sum is over all SNPs in our analysis, and p_{ik} and q_{ik} = 1-p_{ik} are the allele frequencies
406 for the *i*-th SNP. Hardy-Weinberg equilibrium was assumed. Then for every individual *j* in the *k*-
407 th subpopulation, we standardized the observed number of heterozygous genotypes, O_{kj}, by
408 comparing it with the expected number H_k:
409 (2) $\left(O_{kj} - H_k\right)/H_k$
410 Standardized heterozygosity was regressed onto birth year using GAM in all study individuals,
411 and the results were plotted for Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian,
412 and Hispanic or Latino.
413 Pairwise linkage disequilibrium (LD) D' statistics were calculated for all pairs of common
414 (MAF > 0.05) SNPs within 10 megabases (Mb) using Haploview software (Wood, 2011). D'
415 statistics were regressed onto physical distance between SNPs using GAM for distances in the
416 interval from 9-10 Mb for each birth decade.
417 We downloaded a 1% representative sample of individual-level responses to the
418 American Community Survey (ACS) from the Integrated Public Use Microdata Series (IPUMS)

419 (Ruggles, 2014). Race and ethnicity was self-reported in the ACS IPUMS. We regressed the
420 number of major race groups claimed by individuals onto their reported birth year using GAM
421 and frequency weights provided by IPUMS for TN, the South East Central census region, and the
422 entire US. For individual groups, such as "White" and "Black or African American", we plotted
423 all individuals who responded affirmatively to those items; thereby, the samples for the
424 individual race group plots are not independent and overlap at observations where participants
425 claim two or more race groups.
426
427 **Phenotype Classification**
428 Each individual was classified according to 1,645 phenotypes based on the International
429 Classification of Disease, Ninth Revision, Clinical Modification (ICD9) Codes (Denny et al., 2013).
430 Our classification strategy includes all ICD9 features except for procedures. Additionally, the
431 system is hierarchical such that disease subtypes are also classified, such as cardiac arrhythmias
432 are the parent to atrial fibrillation and atrial flutter. Additional phenotypes that are not
433 represented directly in the ICD9 hierarchy are also included, such as inflammatory bowel
434 disease as the parent for Crohn's disease and ulcerative colitis. Diagnoses that were not
435 possible for an individual were set to missing, such as pregnancy for males, or prostate diseases
436 for females. Detailed features of all phenotype algorithms used are available from:
437 http://phewascatalog.org.
438
439 **Clinical Outcomes**

440 For each phenotype, we regressed the binary outcome onto the standardized
441 heterozygosity from equation 2 above, adjusted for birth year and the top 5 principal
442 components of ancestry using logistic regression. We limited analysis to outcomes with 40 or
443 more cases and individuals with at least 2 ICD9 codes. We determined the threshold for
444 statistical significance by Bonferroni correction for the number of analyses where the model
445 converged.
446
447 **Data and Software Availability**
448 The data that support the findings of this study are available from the BioVU resource,
449 but restrictions apply to the availability of these data, which were used under license for the
450 current study, and so are not publicly available. Data are however available upon reasonable
451 request with permission from BioVU.

452 **Acknowledgements**

453 The authors would like to thank J. Henshaw and W. Kunce for useful feedback on the
454 manuscript. The dataset used in the analyses described were obtained from Vanderbilt
455 University Medical Centers BioVU which is supported by institutional funding and by the
456 Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. This work was supported by the
457 National Institutes of Health (grant numbers RC2GM092618, U01HG004603, K23HD000001,
458 R25/T32CA160056, R01LM010685).

459

460 **Declaration of Interests**

461 The authors declare no competing interests.

462

463 **References**

464 Baharian, S., Barakatt, M., Gignoux, C.R., Shringarpure, S., Errington, J., Blot, W.J., Bustamante,
465 C.D., Kenny, E.E., Williams, S.M., Aldrich, M.C., et al. (2016). The great migration and African-
466 American genomic diversity. PLOS Genet. *12*, e1006059.

467 Balloux, F., Amos, W., and Coulson, T. (2004). Does heterozygosity estimate inbreeding in real
468 populations? Mol. Ecol. *13*, 3021–3031.

469 Barnes, P.J. (2008). Immunology of asthma and chronic obstructive pulmonary disease. Nat.
470 Rev. Immunol. *8*, 183–192.

471 Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD
472 and haplotype maps. Bioinformatics *21*, 263–265.

473 Bihlmeyer, N.A., Brody, J.A., Smith, A.V., Lunetta, K.L., Nalls, M., Smith, J.A., Tanaka, T., Davies,
474 G., Yu, L., Mirza, S.S., et al. (2014). Genetic diversity is a predictor of mortality in humans. BMC
475 Genet. *15*, 159.

476 Bittles, A.H., and Black, M.L. (2010). Consanguinity, human evolution, and complex diseases.
477 Proc. Natl. Acad. Sci. *107*, 1779–1786.

478 Brito-Zerón, P., Izmirly, P.M., Ramos-Casals, M., Buyon, J.P., and Khamashta, M.A. (2015). The
479 clinical spectrum of autoimmune congenital heart block. Nat. Rev. Rheumatol. *11*, 301–312.

480 Brito-Zerón, P., Baldini, C., Bootsma, H., Bowman, S.J., Jonsson, R., Mariette, X., Sivils, K.,
481 Theander, E., Tzioufas, A., and Ramos-Casals, M. (2016). Sjögren syndrome. Nat. Rev. Dis. Prim.
482 *2*, 16047.

483 Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M.,
484 Bustamante, C.D., and Ostrer, H. (2010). Colloquium paper: genome-wide patterns of
485 population structure and admixture among Hispanic/Latino populations. Proc. Natl. Acad. Sci.
486 U. S. A. *107 Suppl 2*, 8954–8961.

487 Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D., and Mountain, J.L. (2015). The genetic
488 ancestry of African Americans, Latinos, and European Americans across the United States. Am.
489 J. Hum. Genet. *96*, 37–53.

490 Buyon, J.P., Hiebert, R., Copel, J., Craft, J., Friedman, D., Katholi, M., Lee, L.A., Provost, T.T.,

491 Reichlin, M., Rider, L., et al. (1998). Autoimmune-associated congenital heart block:
492 demographics, mortality, morbidity and recurrence rates obtained from a national neonatal
493 lupus registry. J. Am. Coll. Cardiol. *31*, 1658–1666.

494 Campbell, H., Carothers, A.D., Rudan, I., Hayward, C., Biloglav, Z., Barac, L., Pericic, M.,
495 Janicijevic, B., Smolej-Narancic, N., Polasek, O., et al. (2007). Effects of genome-wide
496 heterozygosity on a range of biomedically relevant human quantitative traits. Hum. Mol. Genet.
497 *16*, 233–241.

498 Campbell, H., Rudan, I., Bittles, A.H., and Wright, A.F. (2009). Human population structure,
499 genome autozygosity and human health. Genome Med. *1*, 91.

500 Chapman, J.R., Nakagawa, S., Coltman, D.W., Slate, J., and Sheldon, B.C. (2009). A quantitative
501 review of heterozygosity-fitness correlations in animal populations. Mol. Ecol. *18*, 2746–2765.

502 David, P. (1998). Heterozygosity–fitness correlations: new perspectives on old problems.
503 Heredity. *80*, 531–537.

504 Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang,
505 D., Masys, D.R., Roden, D.M., and Crawford, D.C. (2010). PheWAS: demonstrating the feasibility
506 of a phenome-wide scan to discover gene-disease associations. Bioinformatics *26*, 1205–1210.

507 Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley,
508 J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide
509 association study of electronic medical record data and genome-wide association study data.
510 Nat. Biotechnol. *31*, 1102–1111.

511 Dickerson, G.E. (1973). Inbreeding and heterosis in animals. J. Anim. Sci. *1973*, 54–77.

512 Dumitrescu, L., Ritchie, M.D., Brown-Gentry, K., Pulley, J.M., Basford, M., Denny, J.C.,
513 Oksenberg, J.R., Roden, D.M., Haines, J.L., and Crawford, D.C. (2010). Assessing the accuracy of
514 observer-reported ancestry in a biorepository linked to electronic medical records. Genet. Med.
515 *12*, 648–650.

516 Durbin, R.M., Altshuler, D.L., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark,
517 A.G., Collins, F.S., De La Vega, F.M., Donnelly, P., et al. (2010). A map of human genome
518 variation from population-scale sequencing. Nature *467*, 1061–1073.

519 Falush, D., Stephens, M., and Pritchard, J.K. Inference of Population Structure Using Multilocus
520 Genotype Data: Linked Loci and Correlated Allele Frequencies.

521 Franceschini, N., Fox, E., Zhang, Z., Edwards, T.L., Nalls, M.A., Sung, Y.J., Tayo, B.O., Sun, Y. V.,
522 Gottesman, O., Adeyemo, A., et al. (2013). Genome-wide association Analysis of blood-pressure
523 traits in African-Ancestry Individuals reveals common associated genes in African and Non-
524 African populations. Am. J. Hum. Genet. *93*, 545–554.

525 Frankham, R., Ballou, J.D., Eldridge, M.D.B., Lacy, R.C., Ralls, K., Dudash, M.R., and Fenster, C.B.
526 (2011). Predicting the probability of outbreeding depression. Conserv. Biol. *25*, 465–475.

527 Galdas, P.M., Cheater, F., and Marshall, P. (2005). Men and health help-seeking behaviour:
528 literature review. J. Adv. Nurs. *49*, 616–623.

529 Han, E., Carbonetto, P., Curtis, R.E., Wang, Y., Granka, J.M., Byrnes, J., Noto, K., Kermany, A.R.,
530 Myres, N.M., Barber, M.J., et al. (2017). Clustering of 770,000 genomes reveals post-colonial

531 population structure of North America. Nat. Commun. *8*, 14238.

532 Head, M.L., Kahn, A.T., Henshaw, J.M., Keogh, J.S., and Jennions, M.D. (2017). Sexual selection
533 on male body size, genital length and heterozygosity: Consistency across habitats and social
534 settings. J. Anim. Ecol. *86*, 1458–1468.

535 Kato, N., Loh, M., Takeuchi, F., Verweij, N., Wang, X., Zhang, W., Kelly, T.N., Saleheen, D., Lehne,
536 B., Leach, I.M., et al. (2015). Trans-ancestry genome-wide association study identifies 12 genetic
537 loci influencing blood pressure and implicates a role for DNA methylation. Nat. Genet. *47*,
538 1282–1293.

539 Keller, L.F., and Waller, D.M. (2002). Inbreeding effects in wild populations. Trends Ecol. Evol.
540 *17*, 230–241.

541 Koch, E., Ristroph, M., and Kirkpatrick, M. (2013). Long range linkage disequilibrium across the
542 human genome. PLoS One *8*, e80754.

543 Kodaman, N., Aldrich, M.C., Smith, J.R., Signorello, L.B., Bradley, K., Breyer, J., Cohen, S.S., Long,
544 J., Cai, Q., Giles, J., et al. (2013). A small number of candidate gene SNPs reveal continental
545 ancestry in african americans. Ann. Hum. Genet. *77*, 56–66.

546 Lamkey, and Edwards, J. (1999). Quantitative genetics of heterosis. Genet. Explor. Heterosis
547 Crop. Am. Soc. Agron. Madison.

548 Lofquist, D., Lugaila, T., O'Connell, M., and Feliz, S. (2012). Household and families: 2010. 2010
549 census briefs.

550 Lynch, M. (1991). The genetic interpretation of inbreeding depression and outbreeding

551 depression. Evolution (N. Y). *45*, 622–629.

552 Maglo, K.N., Mersha, T.B., and Martin, L.J. (2016). Population genomics and the statistical
553 values of race: an interdisciplinary perspective on the biological classification of human
554 populations and implications for clinical genetic epidemiological research. Front. Genet. *7*, 22.

555 Mansfield, A., Addis, M., Health, J.M.-J. of M., and 2003, U. (2003). "Why won't he go to the
556 doctor?": The psychology of men's help seeking. Int. J. Mens. Health *2*, 93–109.

557 Mittermann, I., Aichberger, K.J., Bünder, R., Mothes, N., Renz, H., and Valenta, R. (2004).
558 Autoimmunity and atopic dermatitis. Curr. Opin. Allergy Clin. Immunol. *4*, 367–371.

559 Moreno-Estrada, A., Gignoux, C.R., Fernandez-Lopez, J.C., Zakharia, F., Sikora, M., Contreras, A.
560 V., Acuna-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). The genetics of
561 Mexico recapitulates Native American substructure and affects biomedical traits. Science (80-.
562). *344*, 1280–1285.

563 Nalls, M.A., Simon-Sanchez, J., Gibbs, J.R., Paisan-Ruiz, C., Bras, J.T., Tanaka, T., Matarin, M.,
564 Scholz, S., Weitz, C., Harris, T.B., et al. (2009). Measures of autozygosity in decline:
565 globalization, urbanization, and its implications for medical genetics. PLoS Genet. *5*, e1000415.

566 Nei, M., and Li, W.-H. (1973). Linkage disequilibrium in subdivided populations. Genetics *75*,
567 213–219.

568 Pemberton, T.J., and Szpiech, Z.A. (2018). Relationship between deleterious variation, genomic
569 autozygosity, and disease risk: insights from The 1000 Genomes Project. Am. J. Hum. Genet.
570 *102*, 658–675.

571 Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006).
572 Principal components analysis corrects for stratification in genome-wide association studies.
573 Nat. Genet. *38*, 904–909.

574 Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using
575 Multilocus Genotype Data.

576 Reich, D., and Patterson, N. (2005). Will admixture mapping work to find disease genes? Philos.
577 Trans. R. Soc. Lond. B. Biol. Sci. *360*, 1605–1607.

578 Roden, D., Pulley, J., Basford, M., Bernard, G., Clayton, E., Balser, J., and Masys, D. (2008).
579 Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin.
580 Pharmacol. Ther. *84*, 362–369.

581 Rosenberg, N.A. (2011). A population-genetic perspective on the similarities and differences
582 among worldwide human populations. Hum. Biol. *83*, 659–684.

583 Rudan, I., Carothers, A.D., Polasek, O., Hayward, C., Vitart, V., Biloglav, Z., Kolcic, I., Zgaga, L.,
584 Ivankovic, D., Vorko-Jovic, A., et al. (2008). Quantifying the increase in average human
585 heterozygosity due to urbanisation. Eur. J. Hum. Genet. *16*, 1097–1102.

586 Ruggles, S. (2014). Big microdata for population research. Demography *51*, 287–297.

587 Ruiz-Narvá Ez, E.A., Rosenberg, L., Wise, L.A., Reich, D., and Palmer, J.R. (2011). Practice of
588 epidemiology validation of a small set of ancestral informative markers for control of
589 population admixture in african americans. Am. J. Epidemiol. *173*.

590 Shull, G.H. (1948). What Is “Heterosis”? Genetics *33*, 439–446.

591 Spergel, J.M. (2010). From atopic dermatitis to asthma: the atopic march. Ann. Allergy, Asthma
592 Immunol. *105*, 99–106.

593 Stefanski, L.A., and Boos, D.D. (2002). The Calculus of M-Estimation. The American Statistician.
594 *56*, 29-38.

595 Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green,
596 J., Landray, M., et al. (2015). UK Biobank: an open access resource for identifying the causes of
597 a wide range of complex diseases of middle and old age. PLOS Med. *12*, e1001779.

598 Szulkin, M., Bierne, N., and David, P. (2010). Heterozygosity-fitness correlations: a time for
599 reappraisal. Evolution (N. Y). *64*, 1202–1217.

600 Tedeschi, A., and Asero, R. (2008). Asthma and autoimmunity: a complex but intriguing relation.
601 Expert Rev. Clin. Immunol. *4*, 767–776.

602 Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B.,
603 Awomoyi, A.A., Bodo, J.-M., Doumbo, O., et al. (2009). The genetic structure and history of
604 Africans and African Americans. Science *324*, 1035–1044.

605 Verdu, P., and Rosenberg, N.A. (2011). A general mechanistic model for admixture histories of
606 hybrid populations. Genetics *189*, 1413–1426.

607 Villuendas, R., Olivé, A., Juncà, G., Salvador, I., Martínez-Morillo, M., Santos-Pardo, I.,
608 Pereferrer, D., Zamora, E., and Bayes-Genis, A. (2014). Autoimmunity and Atrioventricular Block
609 of Unknown Etiology in Adults. J. Am. Coll. Cardiol. *63*, 1335–1336.

610 Wang, C., Xu, Z., Jin, G., Hu, Z., Dai, J., Ma, H., Jiang, Y., Hu, L., Chu, M., Cao, S., et al. (2013).

611 Genome-wide analysis of runs of homozygosity identifies new susceptibility regions of lung
612 cancer in Han Chinese. J. Biomed. Res. *27*, 208–214.

613 Wood, S.N. (2010). Fast stable REML and ML estimation of semiparametric GLMs. J. R. Stat. Soc.
614 Ser. B Statistical Methodol. *73*, 3–36.

615