RESEARCH ARTICLE

The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians

Hanxiao Sun[1☯], Meng Lin[1☯], Emily M. Russell[2], Ryan L. Minster[2], Tsz Fung Chan[1], Bryan L. Dinh[1,3], Take Naseri[4], Muagututi‘a Sefuiva Reupena[5], Annette Lum-Jones[6], the Samoan Obesity, Lifestyle, and Genetic Adaptations (OLaGA) Study Group[¶], Iona Cheng[7], Lynne R. Wilkens[6], Loïc Le Marchand[6], Christopher A. Haiman[1], Charleston W. K. Chiang[1,3]*

1 Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, **2** Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **3** Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California, United States of America, **4** Ministry of Health, Government of Samoa, Apia, Samoa, **5** Lutia i Puava ae Mapu i Fagalele, Apia, Samoa, **6** Epidemiology Program, University of Hawai‘i Cancer Center, University of Hawai‘i, Manoa, Honolulu, Hawaii, United States of America, **7** Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, United States of America

☯ These authors contributed equally to this work.
¶ Membership of Samoan Obesity, Lifestyle, and Genetic Adaptations (OLaGA) Study Group is provided in the Acknowledgement.
* charleston.chiang@med.usc.edu

Abstract

Epidemiological studies of obesity, Type-2 diabetes (T2D), cardiovascular diseases and several common cancers have revealed an increased risk in Native Hawaiians compared to European- or Asian-Americans living in the Hawaiian islands. However, there remains a gap in our understanding of the genetic factors that affect the health of Native Hawaiians. To fill this gap, we studied the genetic risk factors at both the chromosomal and sub-chromosomal scales using genome-wide SNP array data on ~4,000 Native Hawaiians from the Multiethnic Cohort. We estimated the genomic proportion of Native Hawaiian ancestry (“global ancestry,” which we presumed to be Polynesian in origin), as well as this ancestral component along each chromosome (“local ancestry”) and tested their respective association with binary and quantitative cardiometabolic traits. After attempting to adjust for non-genetic covariates evaluated through questionnaires, we found that per 10% increase in global Polynesian genetic ancestry, there is a respective 8.6%, and 11.0% increase in the odds of being diabetic ($P = 1.65{\times}10^{-4}$) and having heart failure ($P = 2.18{\times}10^{-4}$), as well as a 0.059 s.d. increase in BMI ($P = 1.04{\times}10^{-10}$). When testing the association of local Polynesian ancestry with risk of disease or biomarkers, we identified a chr6 region associated with T2D. This association was driven by an uniquely prevalent variant in Polynesian ancestry individuals. However, we could not replicate this finding in an independent Polynesian cohort from Samoa due to the small sample size of the replication cohort. In conclusion, we showed that Polynesian ancestry, which likely capture both genetic and lifestyle risk factors, is

OPEN ACCESS

Citation: Sun H, Lin M, Russell EM, Minster RL, Chan TF, Dinh BL, et al. (2021) The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians. PLoS Genet 17(2): e1009273. https://doi.org/10.1371/journal.pgen.1009273

Editor: Joseph Lachance, Georgia Institute of Technology, UNITED STATES

Received: May 12, 2020

Accepted: November 18, 2020

Published: February 11, 2021

Data Availability Statement: All genotype data are available from phs000220.v2.p2 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000220.v2.p2).

Funding: The Multiethnic Cohort was funded through grants from the National Cancer Institute (U01CA164973, P01CA168530) and National Human Genome Research Institute (U01HG007397). Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National

Heart, Lung and Blood Institute (NHLBI). NHLBI TOPMed: Genome-wide Association Study of Adiposity in Samoans was funded by NHLBI (R01HL093093 [S.T. McGarvey] and (R01 HL133040 [R.L.M.]); WGS (phs000972) was performed at the University of Washington Northwest Genomics Center (HHSN268201100037C) and the New York Genome Center (HHSN268201500016C). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I). This research is also supported by Samoan Ministry of Health and the Samoa Bureau of Statistics. Computation for this work is supported by the University of Southern California's Center for Advanced Research Computing (https://carc.usc.edu). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

associated with an increased risk of obesity, Type-2 diabetes, and heart failure, and that larger cohorts of Polynesian ancestry individuals will be needed to replicate the putative association on chr6 with T2D.

Author summary

Native Hawaiians are one of the fastest growing ethnic minorities in the U.S., and exhibit increased risk for metabolic and cardiovascular diseases. However, they are generally understudied, especially from a genetic perspective. To fill this gap, we studied the association of Polynesian genetic ancestry, at genomic and subgenomic scales, with quantitative and binary traits in self-identified Native Hawaiians. We showed that Polynesian ancestry, which likely captures both genetic and non-genetic risk factors related to Native Hawaiian people and culture, is associated with increased risk for obesity, type-2 diabetes, and heart failure. While we do not endorse utilizing genetic information to supplant current standards of defining community membership through self-identity or genealogical records, our results suggest future studies could identify population-specific genetic susceptibility factors that may elucidate underlying biological mechanisms and reducing the disparity in disease risks in Polynesian populations.

Introduction

Native Hawaiians are the second fastest growing ethnic group in the U.S., growing 40% from the 2000 to 2010 U.S. census [1]. Moreover, Native Hawaiians display alarming rates of obesity, coronary heart disease, diabetes, cardiovascular diseases, cancers, and other related chronic health conditions [2–9]. Epidemiological studies have shown that 49% of adult Native Hawaiians are obese, compared to 21% of European Americans and 13% of Japanese Americans living in Hawaiʻi [3], with > 2x and 5x higher odds of being obese than European- and Asian-Americans, respectively, after adjusting for socioeconomic status [6]. In addition, Native Hawaiians are ~2–3 times more likely to develop Type-2 diabetes (T2D) than their European American counterparts, even after adjusting for common modifiable risk factors such as BMI and socioeconomic covariates [4]. Similarly, Native Hawaiians are ~1.7 times more likely to develop cardiovascular diseases than European Americans [8], and cardiometabolic risk factors such as hypertension have been shown to be associated with genealogical estimates of proportion of Native Hawaiian ancestry [9]. Taken together, these observations suggest that in addition to non-genetic risk factors such as lifestyle or diet, there may be systematic differences in the number, frequency, or effect size of genetic risk alleles that contribute to epidemiological differences in disease risks between Native Hawaiians and other continental populations. Yet, such genetic investigation has not been conducted and despite awareness and efforts to include more non-European populations in genomic studies, indigenous populations such as Native Hawaiians remain understudied [10–12].

Today, Native Hawaiians are an admixed population. Their ancestors settled the Hawaiʻi archipelagos approximately 1,200–2,000 years ago and remained isolated there until 1778 when they encountered Western explorers who brought novel infectious agents that decimated the Native Hawaiian population before they rebounded over the last couple of centuries [13–16]. During the 18th and 19th centuries, Native Hawaiians became admixed with European and East Asian immigrants to the islands. The 2010 U.S. census data suggests that only approximately 1.2 million individuals in the U.S. derive some proportion of their ancestry from Native Hawaiians, accounting for about 0.4% of the U.S. population. The small population size may be one of the challenges in recruiting large cohorts, which contributes to the reason that this population is under-investigated from a genetic standpoint.

To begin filling the missing gap in the genetic understanding of disease risks in Native Hawaiians, we first distinguished a Native Hawaiian-specific component of ancestry from other continental ancestries, and tested the association of this global (genomic) ancestry to complex traits and diseases in Native Hawaiians. We presumed this component of ancestry to be Polynesian in origin, although we cannot discount the possibility that this component of ancestry has diverged from the prevalent ancestry component found in other extant Polynesian populations today. We further stress that associations between estimated global Polynesian ancestry and any phenotype will also capture any non-genetic cultural or environmental effects that are correlated with Polynesian ancestry. These variables are typically measured with considerable error; thus, adjustment for them does not necessarily exclude residual effects. Therefore, an observed association with genetic ancestry is not evidence for a deterministic impact attributed to the Polynesian genetic ancestry alone. Nevertheless, an observed association with genetic ancestry may imply that genetic mapping studies could identify genetic susceptibility factors enriched in the Polynesian populations and elucidate underlying biological mechanisms.

We then tested the association of local Polynesian ancestry with complex traits and diseases in what is known as admixture mapping. Admixture mapping assumes that causal variants leading to increased risk or trait values occur more frequently on chromosomal segments inherited from the ancestral population that has higher disease risk or larger average trait values [17–19]. This technique is thus ideal as a first line analysis in understudied populations that are recently admixed. It has previously been used in African-American and Latino populations to identify novel genomic regions associated with phenotypes such as asthma, blood cell traits, breast and prostate cancer (reviewed in ref [17]), but has not yet been applied to Native Hawaiians.

Results

Impact of global genetic ancestry on cardiometabolic traits in Native Hawaiians

We used 3,940 self-identified Native Hawaiians from the Multiethnic Cohort (MEC) [20] that were genotyped on the MEGA array [21] to assess the impact of global ancestry on health. We first needed to construct a reference panel for Polynesian (PNS) ancestry since there is no publicly available reference panel for the PNS ancestry among Native Hawaiians. (Note: we refer to this ancestral component as Polynesian for simplicity.) Among the 3,940 Native Hawaiians in our dataset, we identified a panel of 178 unrelated Native Hawaiian individuals with the highest estimated amount of PNS ancestry (>90% in unsupervised ADMIXTURE analysis; **Methods**) after accounting for other sources of recent admixtures, namely Europeans (EUR), East Asians (EAS), and Africans (AFR). Using this reference panel, we computed a haplotype-based estimate of global genetic ancestry for each of the remaining 3,762 individuals, and kept 3,428 unrelated individuals after excluding for the first-degree relatedness in our dataset (**Methods**).

We then assessed in Native Hawaiians the association of each component of ancestries with a set of quantitative and binary cardiometabolic traits. Specifically, we focused on three disease categories for which the Native Hawaiians have shown increased risks in previous epidemiological studies: obesity [3,6], T2D [4], and cardiovascular disease [8,9]. We also examined quantitative traits and biomarkers associated with these diseases, namely BMI at baseline, fasting glucose and insulin level, HDL, LDL, triglycerides, and total cholesterol. More importantly, because non-genetic factors, such as socioeconomic status (SES) and lifestyle factors, could potentially confound the association between global genetic ancestry and risk of diseases, we attempted to adjust for these factors using education as individual level proxy to SES (**Methods**). Overall, we found that higher PNS ancestry is strongly associated with higher risk of obesity, T2D, heart failure (HF), and consistently, with higher BMI and lower HDL levels among the quantitative traits (**Tables 1 and S1–S15**). The associations are consistent with a broadly linear relationship observed between these traits and proportion of PNS ancestry (**S1 Fig**; though there may be some non-linear relationship for T2D and HF among individuals with higher estimated PNS ancestry). For example, we observed in our statistical model that, holding the proportion of EAS and AFR ancestry constant, every 10% increase in the PNS ancestry in our cohort corresponded to a 0.059 s.d. (or 0.35 BMI unit) increase in BMI and a 1.09 times the odds of T2D (after adjusting for BMI). We observed opposite effects of PNS ancestry on waist-to-hip ratio (WHR) in males and females separately, though the statistical significance is

Table 1. Summary of association between global genetic ancestry and quantitative and binary cardiometabolic traits in the Native Hawaiians.

	PNS		**EAS**		**AFR**	
Trait	**β**	**P-value**	**β**	**P-value**	**β**	**P-value**
Quantitative Traits						
BMI	**0.5923**	**1.04×10^{-10}**	**-0.6400**	$<$**2×10^{-16}**	1.0777	0.0948
WHR (male)	-0.3592	0.0179	-0.1358	0.2664	1.9487	0.1218
WHR (female)	0.2272	0.0985	0.2587	0.0139	-0.1722	0.8524
Glucose	-0.0129	0.929	0.1355	0.232	0.7825	0.463
Insulin	0.2858	0.0472	0.0048	0.966	0.6249	0.5573
HDL	**-0.4715**	**1.40×10^{-4}**	0.1753	0.0700	-1.4498	0.0988
LDL	0.0736	0.557	0.0720	0.463	-0.5192	0.559
TG	0.1387	0.0342	0.1426	0.0053	0.1639	0.7237
TC	-0.0441	0.7228	0.2072	0.0331	-0.8084	0.3602
Categorical Traits						
Obesity	**1.2164**	**2.24×10^{-7}**	**-1.3596**	**5.40×10^{-11}**	1.3181	0.3979
T2D	**1.2416**	**$1.04x10^{-9}$**	**0.6836**	**$2.05x10^{-5}$**	1.1393	0.4220
T2D (adj BMI)	**0.8209**	**1.65×10^{-4}**	**1.1765**	**1.30×10^{-11}**	0.3603	0.8125
HF	**1.3104**	**$1.99x10^{-6}$**	-0.0224	0.9209	3.4587	0.0493
HF (adj BMI)	**1.0465**	**2.18×10^{-4}**	0.2528	0.2797	3.4653	0.0593
HYPERL*	0.0652	0.792	**0.6973**	**5.30×10^{-4}**	0.9841	0.5731
HYPERT	0.6461	0.0100	**0.7363**	**$2.74x10^{-4}$**	-0.6385	0.7081
HYPERT (adj BMI)	0.3842	0.135	**0.8783**	**1.89×10^{-5}**	-0.9459	0.585
IHD	0.4787	0.0496	0.0164	0.9327	-0.3750	0.8270
IHD (adj BMI)	0.2881	0.2457	0.1445	0.4633	-0.7074	0.6871
TIA	0.4876	0.143	-0.0201	0.941	2.3080	0.278
TIA (adj BMI)	0.3612	0.2839	0.0719	0.7928	2.1347	0.3226

For each trait tested as regressand, we present the effect sizes (β, in units of s.d. for quantitative traits and log odds for binary traits) and p-values for the final model after accounting for covariates and jointly modeling PNS, EAS, and AFR ancestries as regressors (**Methods**). In all binary traits other than obesity, results adjusting for BMI as a covariate in the model are also reported (* BMI was not found to be associated with HYPERL and thus was not adjusted in the model). Effect sizes and P-values that are significant after adjusting for testing 14 traits are bolded. Abbreviations: BMI, body mass index; HDL, high-density lipoprotein; LDL, low-density lipoprotein; TG, triglycerides; TC, total cholesterol; T2D, Type-2 diabetes; HF, heart failure; HYPERL, hyperlipidemia; HYPERT, hypertension; IHD, ischemic heart disease; TIA, stroke and transient ischemic attack. For full model of each trait tested, please refer to **S1–S15 Tables**.

https://doi.org/10.1371/journal.pgen.1009273.t001

marginal (**Tables 1 and S2**). For T2D, HF, hypertension (HYPERT), and ischemic heart disease (IHD), BMI is an established risk factor. In our models, we also found BMI to be strongly associated with disease risk for these conditions (max $P < 1\times10^{-7}$; **S10, S11 and S13, S14 Tables**). For T2D and HF, we observed a strong association between disease risk and PNS ancestry even after accounting for BMI, suggesting additional risk factors that are specific or correlated to the PNS ancestry (**Table 1**). For HYPERT and IHD, we observed a weak but nominally significant association between PNS ancestry and disease risk if we do not account for BMI. In fact, for most traits tested, the effect sizes due to PNS ancestry are lower after adjusting for BMI (**S2 Fig**), suggesting that at least part of the excessive risk for these traits may be mediated through BMI.

Other components of ancestry found in Native Hawaiians also exert an effect. For example, we observed that a higher East Asian ancestry component are associated with increased risk of T2D, hyperlipidemia, and hypertension, but lower BMI and lowered risk of obesity (**Table 1**). In some cases, the magnitude of effects due to the EAS component could be larger than the PNS component (e.g. T2D after adjusting for BMI; **Table 1**). Therefore, based on our model at least some of the disease risk or variation in quantitative phenotypes in Native Hawaiians may be partly attributed to or compounded by non-Polynesian components of ancestry (**S3 Fig**).

Because, as mentioned above, non-genetic factors such as socioeconomic status could confound our analysis, we further tested if adding neighborhood SES could account for these associations. Neighborhood SES (nSES) is a validated composite measure created by principal component analysis that incorporates U.S. Census data on education, occupation, unemployment, household income, poverty, rent, and house values [22]. This nSES measure was categorized into quintiles based on the nSES distribution of Hawaii census tracts and Native Hawaiian subjects were assigned a quintile based on their geocoded baseline address (**Methods**). For traits (BMI/obesity, HDL, T2D, and HF) that showed significant association with proportion of PNS ancestry, adding nSES into the model showed that nSES was statistically significantly associated with each outcome, and accounted for some proportion of the risk. However, the association between proportion of PNS ancestry and each of these outcomes remained highly significant, with the exception of HDL, which became nominally significant (**Tables 2** and **S1 and S5 and S9–S11**). We did not observe any interaction between nSES and ancestry (**S16 Table**). These results are again consistent with the possibility that unique Polynesian genetic risk factors exist in the Native Hawaiians that partly explain the elevated risk.

As the strongest association with genetic ancestry came from BMI, we further investigated the association between BMI and PNS ancestry in stratified analysis. We found little evidence

Table 2. Summary of association between global genetic ancestry and quantitative and binary cardiometabolic traits in the Native Hawaiians, after adding nSES into the previous model that only adjusted for inidivdual level covariates.

	PNS		EAS		AFR	
Trait	**β**	**P-value**	**β**	**P-value**	**β**	**P-value**
BMI	**0.4974**	**2.26×10^{-7}**	**-0.6113**	**1.37×10^{-15}**	0.8537	0.201
HDL	-0.3027	0.0201	0.1725	0.0891	-0.9393	0.3021
Obesity	**1.0774**	**1.29×10^{-5}**	**-1.3105**	**1.03×10^{-9}**	1.2973	0.4141
T2D (adj BMI)	**0.7575**	**8.51×10^{-4}**	**1.1569**	**6.92×10^{-11}**	0.2535	0.8668
HF (adj BMI)	**1.0201**	**7.41×10^{-4}**	0.2775	0.2588	3.4056	0.0714

We present the effect sizes (β, in units of s.d. for quantitative traits and log odds for binary traits) and p-values for the final model after accounting for covariates for PNS, EAS, and AFR ancestries, using propotion of EUR ancestry as baseline (**Methods**). Effect sizes and P-values that are significant after adjusting for testing 14 traits are bolded.

https://doi.org/10.1371/journal.pgen.1009273.t002

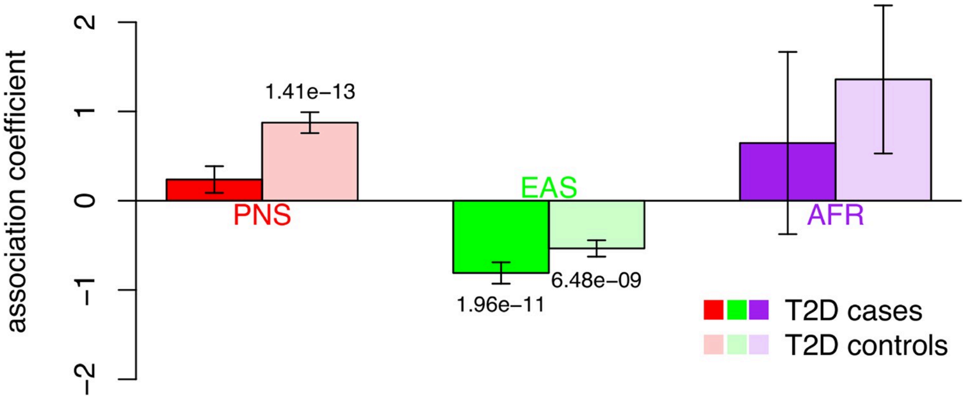

Fig 1. Stratified association testing between global genetic ancestry and BMI. Individuals were stratified based on T2D disease status. Cases are colored in darker color, controls in lighter color. The association coefficient between ancestry component and BMI within strata are shown (**S18** Table). Error bars reflect the estimated standard error from the regression model. P-values for significant association coefficients are labeled. The strongly significant association between PNS ancestry and BMI among T2D controls, but not cases, is suggestive of an interaction between PNS ancestry and T2D.

https://doi.org/10.1371/journal.pgen.1009273.g001

of difference between sexes (**S17 Table**). However, we did observe a strong difference in the strength of association stratified by T2D disease status. Specifically, among T2D cases, we found no significant association between BMI and proportion of PNS ancestry (*P* = 0.112; **Fig 1 and S18 Table**). On the other hand, among T2D controls, individuals were predicted to have 0.087 s.d. (or 0.51 units) or higher BMI per 10% increase in PNS ancestry (*P* = 1.4x10^{-13}). This is despite the T2D strata having similar sample sizes (1,310 cases vs. 1,799 controls). A BMI model including interaction between T2D strata and PNS ancestry showed significant negative interaction (*P* = 0.0004, **S19 Table**).

For a subset of ~300 Native Hawaiians in our cohort, we also have measures of subcutaneous fat and visceral fat, as well as lean mass vs. fat mass obtained through dual-energy x-ray absorptiometry and abdominal magnetic resonance imaging [23]. In this small subcohort, we found that increasing PNS ancestry appears positively associated with subcutaneous fat (*P* = 4.88x10^{-6}) and visceral fat (*P* = 0.014) (**Table 3**). There was no association with lean-to-fat mass ratio (*P* = 0.76), suggesting that PNS ancestry is associated with body fat distribution but not necessarily body fat composition. Because anthropometric measures of body fat distribution such as Waist-to-hip ratio often differ between male and females, we also conducted sex-stratified analysis. We observed similar trend of associations between subcutaneous fat vs. visceral fat, though the association seems more strongly driven by males (**Table 3**).

Mapping of cardiometabolic traits using local genetic ancestry in Native Hawaiians

We next examined the impact of local genetic ancestry on cardiometabolic traits in Native Hawaiians through admixture mapping using linear or logistic regression models. We only analyzed the traits that exhibited a significant association with the global PNS ancestry. We used a threshold of 2.2x10^{-5} to declare genome-wide significance with a trait (**Methods**).

Across the 2 quantitative (BMI and HDL) and 2 binary (T2D and HF; obesity was not included as the definition of obese status is dependent on BMI) traits examined through admixture mapping, we identified one region that surpassed our genome-wide significance threshold (**Fig 2**): 62.7Mb to 65.7Mb on chr6 for T2D (**Table 4 and Fig 2**). We further defined a broader region encompassing neighboring regions with admixture *P*-value less than 1x10^{-4}

Table 3. Association of global ancestry with measures of fat distribution or fat composition among Native Hawaiians.

	Combined		Male		Female	
	Estimate (s.e.)	**P**	**Estimate (s.e.)**	**P**	**Estimate (s.e.)**	**P**
Subcutaneous Fat						
Intercept	-0.39 (0.25)	0.11	-0.58 (0.35)	0.10	-0.25 (0.37)	0.51
PNS	1.30 (0.28)	**4.88×10^{-6}**	1.52 (0.4)	**2.29×10^{-4}**	1.14 (0.4)	**0.0052**
EAS	-0.14 (0.22)	0.54	0.050 (0.33)	0.88	-0.28 (0.32)	0.38
AFR	3.16 (1.80)	0.080	5.45 (3.62)	0.13	2.46 (2.12)	0.25
Total fat mass	-0.0044 (0.0081)	0.59	-0.0033 (0.013)	0.80	-0.0056 (0.011)	0.62
Visceral Fat						
Intercept	-0.26 (0.26)	0.31	-0.11 (0.37)	0.77	-0.48 (0.38)	0.20
PNS	0.71 (0.29)	**0.014**	0.55 (0.42)	0.19	0.98 (0.4)	**0.016**
EAS	-0.29 (0.23)	0.21	-0.092 (0.35)	0.79	-0.36 (0.32)	0.27
AFR	3.45 (1.85)	0.064	6.54 (3.82)	0.089	2.45 (2.13)	0.25
Total fat mass	0.0012 (0.0083)	0.89	-0.0082 (0.013)	0.54	0.0073 (0.011)	0.52
Lean-to-Fat Mass Ratio						
Intercept	0.11 (0.18)	0.55	0.49 (0.28)	0.079	-0.23 (0.25)	0.37
PNS	-0.087 (0.29)	0.76	-0.64 (0.42)	0.13	0.40 (0.39)	0.31
EAS	-0.28 (0.24)	0.24	-0.77 (0.35)	0.030	0.15 (0.33)	0.64
AFR	0.86 (1.95)	0.66	0.55 (4.21)	0.90	1.02 (2.21)	0.64

N = 280, 280, and 294 for analysis on subcutaneous fat, visceral fat, and lean-to-fat mass ratio, respectively. In sex-stratified analysis, N = 128, 128, and 136 males for analysis on subcutaneous fat, visceral fat, and lean-to-fat mass ratio, respectively; N = 152, 152, and 158 females for analysis on subcutaneous fat, visceral fat, and lean-to-fat mass ratio, respectively. Subcutaneous and visceral fats are in units of sample standard deviation (**Methods**)

https://doi.org/10.1371/journal.pgen.1009273.t003

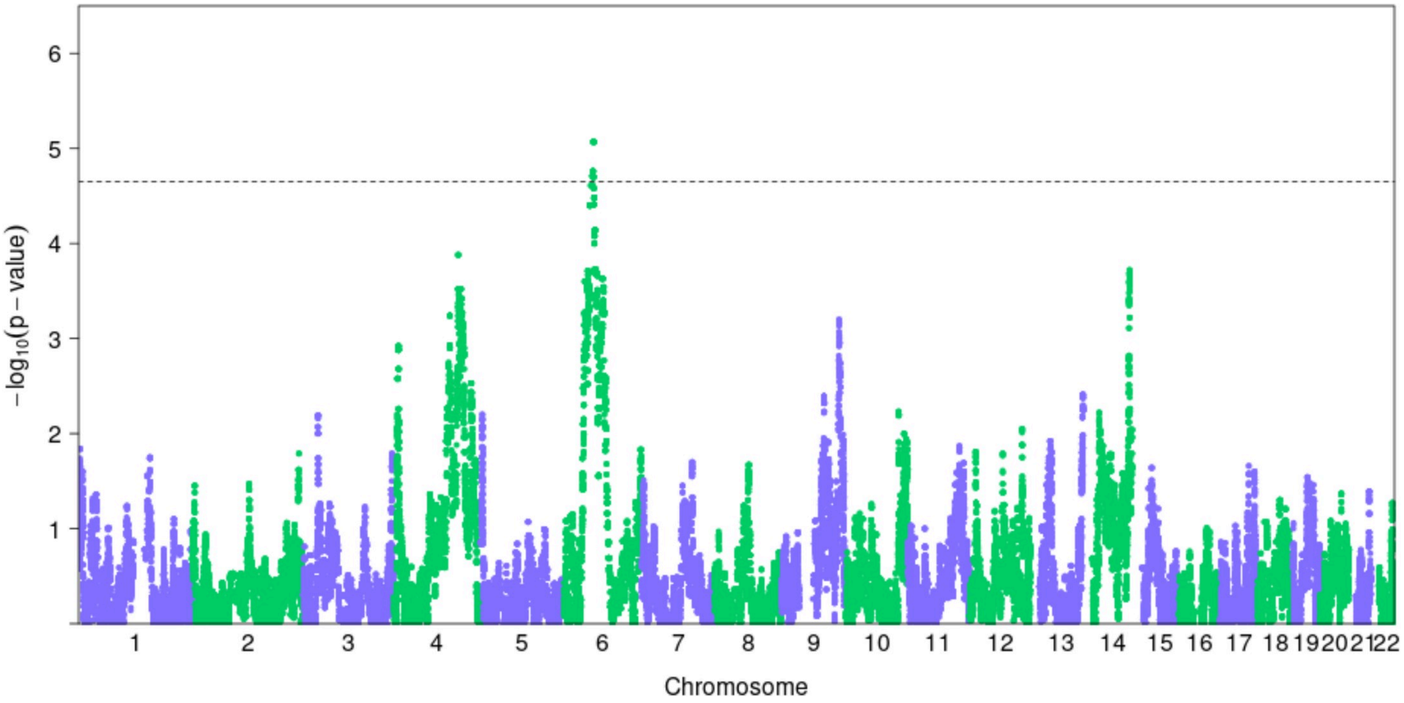

Fig 2. Manhattan plot of admixture mapping results for T2D. Dotted line denotes the genome-wide significance threshold for each trait at 2.2×10^{-5}, determined through permutation.

https://doi.org/10.1371/journal.pgen.1009273.g002

Table 4. Summary of significant loci identified through admixture mapping.

		Admixture Mapping					Single Variant Top Signal				
Trait	**Chr**	**Peak -$\log_{10}$P**	**Signal Region Start–Stop (hg19)**		**Broad Region Start–Stop (hg19)**		**N_{SNP} tested**	**SNP ID**	**OR**	**P-value** *	**Nearest Gene**
T2D	6	5.07	62,697,746	65,763,203	57,098,973	68,542,828	29,751	rs370140172	1.096	1.25×10^{-5}	EYS

The signal region from admixture mapping was defined as the interval between which admixture mapping P-value is below the genome-wide significance threshold of 2.2×10^{-5} ($-\log_{10}P = 4.65$). Broad region was defined as the interval between which admixture mapping P-value is below 1×10^{-4}; continuous segments with admixture mapping P-value below 1×10^{-4} but within 5 Mb are also merged. Within each broad region, we report the variant with strongest association with the trait through either single variant association testing.

* For the chromosome 6 region, rs370140172 would be significantly associated with T2D after correcting for number of variants tested in the region by permutation (at 5% FDR, critical threshold = 1.51×10^{-5}).

https://doi.org/10.1371/journal.pgen.1009273.t004

as potential regions that may harbor causal allele(s). For this broader region spanning 11.4 Mb on chr6 (Table 4), we examined if known variants reported in the GWAS catalog could account for the signals we found through admixture mapping. We found 2 variants in the GWAS catalog for T2D that fall within our admixture peak (S20 Table). We imputed these two variants using 1000Genomes (phase 3) as the reference panel and found that conditioning on these variants did not significantly change our admixture mapping results (top P-value = 6.22×10^{-6}; S4 Fig). These results suggest that our signals detected through admixture mapping may potentially be novel.

To fine-map the candidate region on chromosome 6, we conducted single variant association tests (Fig 3). We imputed the full dataset of 3,940 individuals using 1000 Genomes as reference to increase coverage across the region, and accounted for cryptic relatedness and population structure in a logistic mixed model (**Methods**). We found that the top associated variant on chr6 for T2D was a well-imputed (INFO score = 0.86) 5' UTR variant rs370140172 (OR = 1.096, $P = 1.25 \times 10^{-5}$, Fig 3). Its association with T2D was significant after accounting for number of markers tested in this region (regional significance threshold = 1.51×10^{-5}; Table 3). This variant showed a large difference in frequency between Native Hawaiians (MAF = 24.2% among our reference PNS individuals; 11.2% among MEC-NH population) and European (0%) or East Asian (0.9%) individuals from 1000 Genomes, or Oceania (2.0%) and Southeast Asia (1.2%) individuals from the GenomeAsia Pilot Project [24] (S21 Table). Conditioning on rs370140172 also drastically reduced the admixture association signal (minimum $P \sim 0.001$ in the region; S5 Fig). Taken together, these observations suggest that rs370140172 or its proxy could be the allelic association driving the admixture signal.

We attempted to replicate the single variant signal on chromosome 6 by examining the lead variant and its proxies for association with T2D in another Polynesian population of 2,852 Samoan individuals, including 475 cases and 2,377 controls. We found no significant associations (minimum $P = 0.743$; S22 Table). However, we noted that the derived allele of rs370140172 had a significantly lower frequency in the Samoans (8.7%, compared to 24.2% in reference PNS individuals). The lower frequency of the allele and the current sample size does not provide sufficient power to replicate the association signal even at nominal significance level of 0.05 (Power = 14.2%; S6 Fig). Because of the difference in frequency between Native Hawaiians and Samoans, we also examined if this locus exhibits signals of positive natural selection in the Native Hawaiians (**Methods**). We observed that the derived allele is found on the longer haplotypes in Native Hawaiians, although not statistically significantly different from other variants of similar derived allele frequency in the genome (Z-score = +1.31; empirical $P = 0.067$. S7 Fig). Similarly, the allele frequency difference between MEC-NH and the

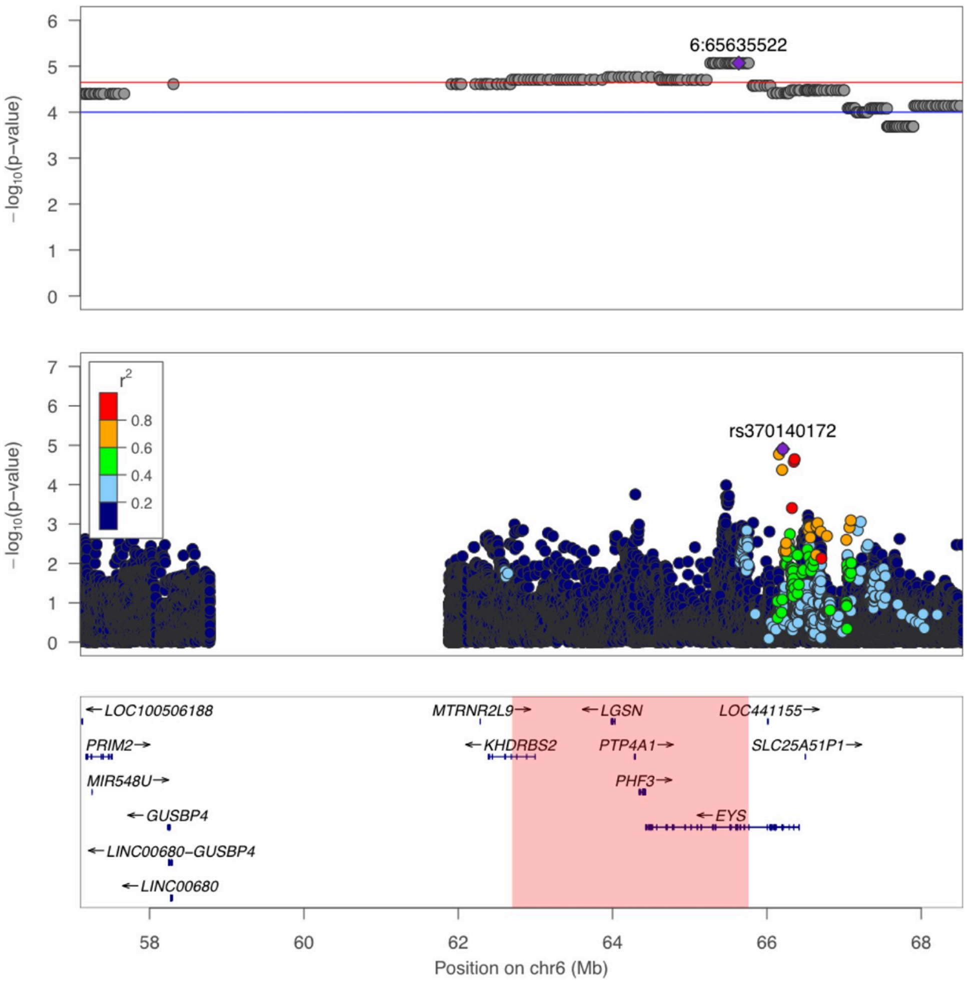

Fig 3. Association signals with T2D in the broad region of chr 6. The top panel depicts association signals from admixture mapping on 3,428 unrelated individuals. The middle panel depicts the single variant association result of the same region, using linear mixed model on all 3,940 individuals with genotype dosages imputed from 1000 Genomes Project. The bottom panel shows the corresponding genomic coordinates and nearby genes. Highlighted region at the bottom indicates the signal region as defined in Table 2. LD between rs370140172 and other variants in the region, as shown in the color legend, was r2 calculated in sample using the imputed genotypes.

https://doi.org/10.1371/journal.pgen.1009273.g003

Samoans for the derived allele at rs370140172 is also marginally insignificant compared to other variants of similar derived allele frequency (AF diff = 15.4%; empirical $P = 0.076$. S7 Fig). Thus, the elevated frequency in MEC-NH may still be the result of genetic drift.

Discussion

This study aimed to fill a gap of genetic research in Native Hawaiians. We focused on studying the association of genetic ancestry, both globally and locally, to diseases for which Native Hawaiians showed increased risk. While the focus is on genetic ancestry, we emphasize that our approach does not constitute a methodology to quantify the degree of indigenousness among individuals native to the Hawaiian archipelago. Estimating proportion of genetic ancestry is not without errors. The estimates may change depending on the input sample size, or the choice of the reference data or SNPs used for analysis, and there is a conceptual difference between genetic ancestry and genealogical ancestry. Moreover, there are also difficulties in interpreting the estimated proportion. In this paper we made the simplifying assumption that the predominant component of ancestry found in MEC-NH individuals but not in other continental populations is Polynesian in origin; this does not take into account any post-divergence differentiation between Native Hawaiians and other Polynesian populations. Given these caveats, we therefore believe the approach described here should not supplant current approaches, such as through self-reports or genealogical records, to define community membership. Consistent with this belief, we analyzed all individuals with available genetic data who self-identify as at least part Native Hawaiian ancestry; we did not attempt to define a population of Native Hawaiians using genetic data.

We began our analysis by modeling Polynesian ancestry. We first conducted ADMIXTURE analysis to identify an internal Native Hawaiian ancestry reference panel since there is no appropriate representative panel currently available. Consistent with their known history, we found Native Hawaiians to be a recently admixed population, deriving the largest proportion of their genetic ancestry from a presumed Polynesian ancestral component (on average ~40.2%). We also found that global Polynesian ancestry from MEC-NH is statistically significantly associated with BMI, HDL, Type-2 diabetes, obesity and heart failure after adjusting for other components of ancestries and available non-genetic covariates (**Tables 1 and 2**). Polynesian ancestry was also nominally associated with WHR (in males), insulin level, triglycerides, hypertension and ischemic heart diseases, but these associations did not remain statistically significant after Bonferroni correction for the multiple traits that we tested in this study (**Table 1**).

The strong associations between global ancestry and disease risks or related quantitative phenotypes suggest the presence of population-specific variants that could contribute to the increased risk observed in these populations. For example, a recently reported, Polynesian-specific, *CREBRF* variant discovered in Samoans was strongly associated with the odds of obesity, a finding that we previously replicated in Native Hawaiians [25]. However, we should also stress that an association with global ancestry would also in theory capture any non-genetic cultural or environmental effects that are correlated with ancestry. We attempted to account for non-genetic factors such as individual-level socioeconomic status (SES) like education, behavioral traits such as cigarette smoking, and neighborhood-level composite SES measures based on census data that captures several key domains of SES, including income, poverty, occupation, housing, and others [22,26] (**Table 2**). Admittedly, these variables are still imperfect proxies for SES. For example, while our composite nSES index has been shown to uncover important SES associations for complex traits [27,28], it may overlook specific SES factors that are particularly relevant for Native Hawaiians. Given that non-genetic factors certainly play a role in the etiology of these traits, future studies may further explore both individual-level (*e.g.* physical activity, diet, alcohol or medication use) and domain-focused neighborhood-level (*e.g.* poverty, discrimination) non-genetic factors. Therefore we should interpret these associations with global ancestry with caution.

One notable observation is the association between global PNS ancestry and BMI. In analysis stratified by T2D status, despite having similar numbers of cases and controls, we found that PNS ancestry is not associated with BMI among T2D cases, but is associated with higher BMI among individuals unaffected by T2D (**S18 Table**). In models including interaction between global ancestry and T2D, we again observed that while T2D cases generally have higher BMI, those with greater PNS ancestry would actually have lowered BMI than those with less PNS ancestry (**S19 Table and Fig 1**). EAS ancestry is also strongly associated with BMI in Native Hawaiians, but we found no evidence of interaction between EAS and T2D (**S19 Table**). We explored whether this interaction could be mediated through differential body composition. For example, individuals with increasing PNS ancestry may possess more lean mass, which contributes to BMI, than fat mass, which contributes to BMI and risk for T2D. There are some suggestions that individuals of PNS ancestry preferentially have greater lean mass than fat mass [29], although data is limited and we found no association between PNS ancestry and lean-to-fat mass ratio in a small subcohort of MEC-NH (**Table 2**). We also considered fat distribution. For example, individuals with increasing PNS ancestry may preferentially store fat subcutaneously, which contribute to general adiposity and BMI but not necessarily to T2D, rather than viscerally, which could lead to insulin resistance and contribute to peripheral insulin sensitivity and further T2D [30,31]. We found PNS ancestry to be significantly associated with both subcutaneous fat and visceral fat among our small sub-cohort (**Table 2**), which made it difficult to assess whether there is difference between two types of fat storage. As global ancestry can capture correlated cultural or environment factors in addition to the genetics, perhaps more plausibly, the interaction may signal behavioral modifications correlated with T2D risks. For example, if individuals with greater PNS ancestry are more susceptible to T2D due to a modifiable risk factor that contributes to BMI, but could remove the risk factor after being diagnosed with T2D, there could be a reduced association between BMI and PNS ancestry among T2D cases. Ultimately, more data will be needed to make firm conclusions.

These caveats nonewithstanding, we conducted admixture mapping to test the association between local PNS ancestry genome-wide with the traits significantly associated with global PNS ancestry. Because admixture mapping had not been previously conducted among Native Hawaiians and the haplotypic pattern and LD structure within Native Hawaiians had not been previously explored, we used permutation as well as a recently published method [32] to establish the genome-wide threshold for significance for a single trait, which we determined to be $2.2x10^{-5}$. Using this threshold, we found one notable region on chr6 associated with T2D (**Figs 2** and **3**).

Single variant fine-mapping of the chr6 region identified a significant association on chr6 (rs370140172, $P = 1.25x10^{-5}$) after correcting for number of variants tested by permutation (regional significance threshold = $1.51x10^{-5}$). This variant was imputed with high accuracy (INFO score = 0.86), exhibits large frequency enrichment compared to other populations (24% in non-admixed Native Hawaiians but monomorphic in 1KGP EUR and < 1% in 1KGP EAS; gnomAD v3 overall frequency = 0.00054), and explained the admixture mapping signal we detected (**S5 Fig**). Rs370140172 falls within the 5' UTR (2^{nd} exon) of the gene *EYS*. Because the genetic risk of T2D has been suggested to be mediated through alleles affecting enhancer activity in pancreatic islet cells, we also examined published Hi-C data [33] of this region. We found that rs370140172 did not form a detectable chromatin loop in this dataset, but five subregions within our signal region identified by admixture mapping (chr6:62.7Mb-65.8Mb; **Table 4**) were found to form potential enhancer-promoter loops. In all five cases, the target anchor points are relatively nearby, falling between chr6:64.3Mb-65.0Mb within the *EYS* gene. Mutations in *EYS* are known to cause recessive retinitis pigmentosa [34,35], though there was no obvious link to T2D other than a suggestive association with T2D in Europeans (rs10498828, ~670kb away, $P = 9x10^{-6}$), and a genome-wide association with BMI in a Japanese population (rs148546399, ~1.5Mb away, $P = 1x10^{-9}$) [36,37]. Taken together, rs370140172 or its proxy may signal a novel population-specific candidate locus associated with T2D and, if its effect is mediated through a regulatory mechanism, *EYS* is a plausible causal gene. However, we failed to replicate the association signal at this variant in a Samoan cohort. The failure to replicate could be due to decreased power as the variant has lower frequency in Samoans and the cohort is relatively small. Furthermore, we may be limited by the availability of imputation panels; we used the 1000 Genome Project as reference panel and, as such, a number of Polynesian-specific variants that could underlie the admixture signal in these region may not be well imputed. Improvements in the genomic resources available for studying Polynesian populations as well as increased sample size from these communities will help validate and delineate the observed association at this locus.

Finally, it should be noted that Polynesian ancestry component is not the only component conferring risk to disease in the Native Hawaiian people. In our analysis other components of ancestry could also add to or drive disease risk in Native Hawaiians. For example, we find that T2D risk is independently positively correlated with the estimated proportion of East Asian ancestry in the Native Hawaiians, particularly after adjusting for BMI (**Table 1**). This finding is consistent with previous epidemiological observations based on self-reported recent admixture in the parental generation [38], and may suggest that for these traits admixture mapping focusing on the non-Polynesian components of ancestry may complement large-scale GWAS in these major continental populations.

In summary, Native Hawaiians exhibit an increased risk for obesity, type-2 diabetes, and a number of cardiovascular diseases, but are generally understudied from a genetic standpoint in the literature. We present the first analysis of the genetic ancestry of present-day self-identified Native Hawaiians and suggest that it may have an impact on the risk of these diseases. However, genetic ancestry also reflects non-genetic cultural or environmental effects and we cannot exclude residual confounding by these variables. Nevertheless, a better understanding of the genetic susceptibility risk factors will complement other epidemiological, non-genetic, risk factors for uniquely prevalent diseases among the Native Hawaiians. It is by integrating both genetic and non-genetic risk factors in our understanding of population-specific disease risk that we will have a better chance in clarifying the underlying biological mechanisms and control these diseases. Native Hawaiians have experienced a unique evolutionary history in their trans-Pacific voyages and settlement of the Hawaiian archipelago. Both the demographic and adaptive histories of these people may have shaped their genetic architecture. Further studies focusing on indigenous Polynesian populations, such as Native Hawaiians, will advance the findings reported here and may help alleviate the disparity in genomic medical research existing for Native Hawaiians.

Methods

Study population

In this study, we used genetic and epidemiologic data from Native Hawaiian individuals from the Multiethnic Cohort (MEC). MEC is a prospective epidemiological cohort of >215,000 individuals spanning five major ethnicities, including biospecimen samples on >5,300 Native Hawaiians. It is currently the largest single cohort with genetic information on Native Hawaiians, and thus is ideal for our study. In this study, we used a subcohort of >3,900 individuals genotyped on the MEGA genotyping array [21] as part of the PAGE consortium [39]. The institutional review boards of the University of Hawai'i and the University of Southern California approved the study protocol. All participants signed an informed consent form.

Quality control of MEGA array was previously described [25]. In general, individual and genotype level quality control filters were previously applied as part of PAGE, and additionally we applied the following steps: All variant names were updated to dbSNP v144; duplicated loci and indels were removed; triallelic variants or variants with non-matching alleles to 1000 Genomes Project phase 3 (1KGP) were discarded; loci with unique positions not found in 1KGP were removed from the dataset; alleles were standardized to the positive strand by comparing to 1KGP. Finally, a genotype missingness filter of 5% and a minor allele frequency filter of 1% were applied, resulting in a total of 3,940 MEC Native Hawaiian (MEC-NH) individuals genotyped at 697,505 SNPs.

Global and local ancestry inference

In addition to a predominant Polynesian (PNS) ancestry, Native Hawaiians are known to be recently admixed with individuals of European and East Asian ancestry [14]. In order to define individual genetic ancestry, whether locally or globally, we needed a reference panel for the Polynesian component of the Native Hawaiian ancestry. As such a reference panel does not exist, we sought to construct an internal reference panel by identifying MEC-NH individuals with the largest amount of global Polynesian ancestry as previously described [25]. Briefly, we combined all MEC individuals genotyped on the same MEGA array (3,940 Native Hawaiians, 3,465 Japanese, 30 Hispanic/Latinos, 5,325 African Americans) and all individuals from 1000 genomes Project, pruned SNPs with $r^2 > 0.1$ (using window sizes of 50 SNPs with steps of 10 SNPs across the genome), and partitioned the samples to two groups of related (up to and including 2nd degree) and unrelated individuals by KING (default threshold used). We then ran ADMIXTURE (v. 1.3.0) in unsupervised mode for unrelated samples, then projected the estimated ancestral allele frequency to the related samples to infer the genomic ancestries of the related group. We performed ADMIXTURE analysis in five independent runs, retaining the replicate with the highest likelihood estimated by the software. We found MEC-NH individuals at k = 4 exhibited known components of ancestry from European, East Asian and African, as well as a component of ancestry that is unique to the MEC-NH, presumed to be Polynesian (**S8 Fig**); repeated ADMIXTURE analysis from k = 4 to k = 8 showed this Polynesian component to be stable (**S9 Fig**). We then identified 178 unrelated MEC-NH individuals (kinship coefficient <0.2 estimated from PC-relate [40]) whose Polynesian component of ancestry were estimated to be over 0.9 at k = 4 analysis as reference for the Polynesian component of the Native Hawaiian ancestry.

To call local ancestry, we merged the MEC-NH samples with the above 1000 Genomes reference individuals, and rephrased the merged dataset using EAGLE2 (v 2.4.1). Next, we combined the 178 MEC-NH reference with the above 1KGP reference individuals to form the reference panel. Using this reference panel, we then inferred local ancestry used RFMix [41] (version 2.03-r0). One key parameter for RFMix is the local recombination rates, which vary across continental populations [42,43] but has not been estimated for Native Hawaiians or Polynesians. However, using multi-way admixed 1KGP American (AMR) populations, we evaluated the impact of misspecification of a recombination map. We found that RFMix inferences of local ancestry are robust even using a constant recombination map (>98% concordance, **S23 Table**). Therefore we used HapMap2 pooled recombination map (ftp: //ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/working/20110106_recombination_hotspots/) to infer local ancestry in Native Hawaiians. To obtain global ancestry estimates, we summed the local ancestry estimates across the genome, after excluding tracts that have any ancestral probability < 0.9. We observed that on average, a self-reported Native Hawaiian individual derived ~29.6% ancestry from EUR, ~29.0% ancestry from EAS, ~1.2% ancestry from AFR, and the remaining ~40.2% ancestry from PNS. These values are similar to previous estimates of proportions of genetic ancestry from MEC using ancestry informative markers [44]. The summed PNS ancestry from RFMix is highly concordant with that inferred from ADMIXTURE [25], and is thus used for phenotype association and covariate adjustments in admixture mapping (below).

Phenotype analyzed

We focused on three categories of traits for which the Native Hawaiians exhibit excess risk in past epidemiological studies [2–4,6–9]: (1) adiposity traits, which include BMI at baseline and obesity; (2) metabolic traits, which include fasting glucose level, fasting insulin level, and Type-

2 diabetes (T2D); and (3) cardiovascular traits, which include HDL, LDL, triglycerides (TG), total cholesterol (TC), heart failure (HF), hyperlipidemia (HYPERL), hypertension (HYPERT), ischemic heart disease (IHD), and stroke and transient ischemic attacks (TIA). Fasting glucose and insulin levels were collected after entry to the MEC, between 2001–2006. Obesity, T2D, HF, HYPERL, HYPERT, IHD, and TIA are binary disease outcomes. The metabolic and the quantitative cardiovascular traits were previously studied by PAGE consortium; we thus followed the inclusion criteria and phenotype transformation (based on medication use) as previously suggested by PAGE [39] (**S24 Table**). At a given BMI, Polynesians have a higher proportion of lean muscle mass to fat mass than Europeans so we use the recommended BMI cut-off of 32 kg/m^2 to define obesity cases and controls [29,45]. T2D includes prevalent cases at cohort entry and incident cases during follow-up, based on self-report with medication use in questionnaires or a report from linkage to Hawai'i insurers, CMS, or CHDD [39]. For incident binary cardiovascular traits we utilized the Medicare fee-for-service linkage data for MEC [46] defined as https://www2.ccwdata.org/web/guest/condition-categories. Descriptive summaries of the traits and covariates can be found in **S25 Table**.

For a subset of up to 307 individuals we also analzyed subcutaneous fat, visceral fat, and lean-to-fat mass ratio obtained from whole-body dual-energy X-ray absorptimoetry (DXA) and abdominal magnetic resonance imaging (MRI) scans. For subcutaneous fat and visceral fat, sex-stratified and age-adjusted residuals were standardized. Lean-to-fat mass ratio were calculated by taking the ratio of sex-stratified, age-adjusted standardized measure of total lean mass and total fat mass.

Associations between binary and quantitative traits with global ancestries

We tested the association of global Polynesian ancestry with quantitative and binary traits using linear and logistic regressions, respectively. We focused on the 3428 unrelated individuals after removing first degree relatives determined by KING [47] and individuals used in the internal PNS reference panel. To account for the impact of non-genetic factors that can confound the association between traits and genetic ancestry, covariate-adjusted outcomes were create by regressing out the impact of the non-genetic factors. These include traits such as smoking and education, as proxies for socioeconomic status. For quantitative traits, we first conducted univariate regression of the trait of interest on the non-genetic covariates. We then retained age and sex in the model, as well as all covariates that are nominally significantly associated with the trait. For categorical covariates retained in this procedure, we grouped the non-significantly associated level to reduce the variable down to a ternary or binary variable. We then model the covariates jointly in a multivariate regression model, and then standarized the residuals from this model. The standardized residuals were then used in a multivariate regression model with estimated global Polynesian, East Asian and African ancestries as independent variables, leaving European as the reference. For binary traits, we maintained the same structure, first removing uncorrelated covariates based on univariate logistic regression models. The remaining covariates are then used in a multivariate logistic regression with the addition of global ancestry estimates. The coefficients and p-values associated with the non-genetic covariates and global ancestries from the multivariate regression model are provided in **S1–S15 Tables**.

Adjusting for neighborhood socioeconomic status (nSES) measures

To further assess if the association between global ancestry and outcome could be explained by uncaptured non-genetic factors, we included the nSES variable in our regression models [22]. We determined nSES by subjects' residential census tract using an index derived from principal components of indicator variables of SES (education level; proportion unemployed and with blue collar job; proportion <200% poverty line; proportion employed; median household income, rent and home value) based on 1990 Hawaii Census data. Each Native Hawaiian geocoded baseline address (1993–1996) was assigned a nSES qunitle based on the distribution of neighborhood SES across all census tracts in Hawaii. For traits that showed strong association in **Table 1** (*i.e.* BMI/obesity, HDL, T2D, and HF), we added nSES in the model to account for confounding in the assessment of the association with global ancestry. Because of the area-based design, Native Hawaiian participants residing in the same census tract were assigned the same nSES measure. We thus used a mixed effect model to account for this spatial clustering by including the census tract ID as random effect. We used *lmer* and *glmer* (version 1.1–21) function in R (version 3.6.2) with default parameters.

Mapping of binary and quantitative traits using local Polynesian ancestry

To identify local genomic segments in which the Polynesian ancestry is associated with a trait of interest, we conducted admixture mapping. We focused on the same 3428 unrelated individuals used in global ancestry analysis (above). We used linear or logistic regression to test the association of estimated dosage of Polynesian ancestry from RFMix at each genomic location, while controlling for estimated global ancestry from EUR, EAS, and AFR. Traits were modeled in the same way as above in the global ancestry analysis, except we focused only on individual-level covariates for computational efficiency of genome-wide testing.

We determined the significance threshold for admixture mapping for a given trait using two approaches: by a recently published simulation-based approach [32] and by permutation. For the simulation-based approach, because we were only interested in testing the association of Polynesian ancestry to a trait, we dichotomized estimated local ancestry into Polynesian and non-Polynesian segments to estimate the covariance in local ancestry across the genome. We then estimated the genome-wide significance threshold in admixture mapping to be 2.28x10^{-5} using 10,000 simulations in STEAM [32]. For permutation-based approach, we simulated 1,000 runs of genome-wide admixture mapping, each based on a random phenotype drawn from a standard normal distribution. We then examined the distribution of the most significantly associated p-value from each of the simulations and set at the 5% false discovery level to the threshold of 2.24x10^{-5}. The two thresholds are nearly identical (**S10 Fig**), and are similar to previously suggested threshold among Latinos [48] (4.8x10^{-5}). We thus used 2.2x10^{-5} as the genome-wide significance threshold for admixture.

Conditional analysis and single variant tests in associated admixture region

For the locus we identified through local ancestry association (**Table 4**), we defined the signal region as contiguous variants with admixture *P*-values lower than the genome-wide significance threshold (2.2x10^{-5}, or -log$_{10}$P>4.64). We then defined a broad region by extending the signal region to nearby flanking regions that are (1) < 5Mbp away upstream or downstream from the signal region, and (2) with -log$_{10}$P>4. We then imputed our rephased dataset using Sanger Imputation Service (https://imputation.sanger.ac.uk/). We used 1KGP as the reference panel, and PBWT as the imputation software. We subsequently filtered out indels and loci with low imputation quality (INFO score <0.4), and applied a minor allele frequency filter of 1%. We then investigated whether a previously known variant from the GWAS catalog [49] for the same trait could drive this signal by including all GWAS catalog variants residing in the broad region and passed quality control in our study as covariates in a conditional regression analysis.

We also conducted single variant association based on imputed dosages in the entire broad region. We included all 3,940 samples in this analysis, and corrected the relatedness by using a linear mixed model from EMMAX [50]. The inter-sample relatedness was calculated from PC-relate [40] to account for possible population structure. We followed the same covariate model and phenotype transformation as was done in admixture mapping, except for using the top 10 principal components (PCs) from PC-air [51] as substitutes for the global ancestry covariates. 1,000 permutations were carried out to estimate the regional critical values for significance.

Replication analysis in Samoans

We attempted to replicate the association of rs370140172 and nine other proxies showing the strongest single-variant associations with a cross-sectional population based study of Samoans recruited from Independent Samoa in 2010 [45,52]. This study was approved by the institutional review board of Brown University and the Health Research Committee of the Samoa Ministry of Health. All participants gave written informed consent via consent forms in Samoan language.

The Samoan participants from 2010 were genotyped genome-wide with Affymetrix 6.0 genotyping arrays [45]. A subset of 1,284 Samoan participants were whole-genome sequenced as part of the Trans-Omics for Precision Medicine (TOPMed) Program sponsored by the National Institutes of Health (NIH) National Heart, Lung, and Blood Institute (NHLBI). The sequences were used to produce a reference panel for genotype imputation. Genotypes absent from the Affymetrix genotyping array and present in the reference panel were imputed in the remaining Samoan participants. T2D case and control exclusion criteria were defined to mirror that used in the MEC-NH analyses. Specifically, we removed cases who were pregnant, diagnosed with type 1 diabetes, or under 20 years old. We removed controls with fasting glucose greater than 7 mmol/L. This resulted in 475 cases and 2,377 controls. Association testing was conducted using logistic mixed model regression implemented in lme4qtl [53]. Empirical kinship as estimated from the genotypes was included as a random effect covariate. Age, BMI, education (coded as a continuous variable in six levels), and the first ten PCs were included as fixed effect covariates in the logistic mixed model regression.

The Power of the replication analysis was conducted using the Genetic Association Study power calculator (http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/), assuming the sample size, estimated frequency of rs370140172, a prevalence rate of T2D of 17.1% [52] in Samoans, and a significance threshold of 0.05.

Test of natural selection

We calculated the nSL score [54] of derived alleles across all imputed loci using Selscan [55], after the post imputation quality control. We calculated nSL among 178 MEC-NH reference individuals who had estimated PNS ancestry > 90%, and compared the nSL value for rs370140172 (derived allele of 0.24, and INFO score of 0.87) to that of 44,266 variants selected from the genome matched by imputation uncertainty (INFO score 0.77–0.97) and derived allele frequency (0.23–0.25).

Supporting information

S1 Fig. Relationship between complex traits and proportion of PNS ancestry. For each of five traits that we found significant association with PNS ancestry, we show the mean and standard error of untransformed trait value (for quantitative trait) or proportion of affected (for dichotomous) trait as function of bins of PNS ancestry. The sample size for individual available is given above each bin.
(TIF)

S2 Fig. Correlation of effect sizes attributed to PNS ancestry in the regression model with or without adjustment for BMI. Across the binary traits tested, even if the effect attributable to PNS ancestry is not significant, the effect sizes are lowered if accounting for BMI, suggesting at least part of the excess risk for these traits among Native Hawaiians are mediated through higher BMI associated with the ancestry. Hyperlipidemia was excluded because BMI is not associated with the disease risk in univariate regression model. HF, heart failure; HYPERT, hypertension; IHD, ischemic heart disease; T2D, type-2 diabetes; TIA, stroke and transient ischemic attack.
(TIF)

S3 Fig. Ternary plot of predicted trait value or probability of disease as function of major component of ancestry in Native Hawaiians. For the seven traits in Table 1 in which as least one component of ancestry showed significant association, we estimated the fitted trait value (in units of s.d. for quantitative traits BMI and HDL) or probability of being affected (for dichotomous traits Obesity, T2D, HF, HYPERL, and HYPERT) for each person in our dataset, given their estimated proportion of ancestry. For simplicity, we removed individuals with estimated AFR ancestry > 0.05, and re-normalized the EAS, EUR, and PNS ancestry to sum to 1 in the remaining individual. We used Model 2 in **S1 Table** for BMI and **S5 Table** for HDL to obtain predicted phenotype residual in units of standard deviation. For dichotomous traits, we used Model 2 in **S9**–**S13 Tables** for Obesity, T2D, HF, HYPERL, and HYPERT, respectively, and converted the fitted log-odds to probability of being affected. Because the covariates are included in the model rather than being regressed out in quantitative traits, we assumed fixed values of age = 50, BMI = 30 (except for obesity), sex = male, and education level = college graduates. Contour plots are shown labeling in each plot to display the fitted trait value or probability of being affected.
(TIF)

S4 Fig. Admixture mapping P-value with or without conditioning on variants previously reported in GWAS catalog to be associated with T2D (rs79976124 and rs10498828). Green and blue colors denote SNP level P-value in association testing with and without, respectively, conditioning on known GWAS variants.
(TIF)

S5 Fig. Admixture mapping P-value after conditioning on the most strongly associated variant in single variant analysis in chr6 (T2D) broad region. The originally reported admixture signal (blue) can be explained by the conditioned variant (green), suggesting that these single variants might be novel variants associated with these traits.
(TIF)

S6 Fig. Power of replicating the top signal from single variant analysis with T2D in Samoans. We estimated the power to replicate the top signal (rs370140172) from single variant analysis with T2D in the Samoan cohort, using GAS power calculator (http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html). The prevalence rate of T2D in Samoans set as 17.1%, which was the value averaged over the reported values in both sex [1]. The number of cases (N = 475) and controls (N = 2377) were set to the observed sample size in Samoans. The genotype relative risk was set to estimated OR (1.096) from MEC-NH.
(TIF)

S7 Fig. Inconclusive evidence of selection at rs370140172 by nSL and difference of the derived allele frequency with Samoans. We selected ~44,000 variants across the genome from imputed data matched to rs370140172 by derived allele frequency in MEC-NH and by

imputation quality (**Methods**). We compared the nSL statistics (left) and the difference in frequency of the derived allele between MEC-NH and Samoans (right) at rs370140172 (denoted by the blue vertical line) to the null distribution based on the ~44,000 matched SNP. The evidence of selection for rs370140172 is marginally insignificant by either haplotype length ($P = 0.067$) or allele frequency differentiation ($P = 0.076$).
(TIF)

S8 Fig. Global ancestry proportion estimated from unsupervised ADMIXTURE analysis, after integrating runs of relatedness and unrelatedness. 3,465 MEC Japanese (MEC-JA), 30 MEC Latinos (MEC-LA), 5,325 MEC African Americans (MEC-AA), and 3,940 MEC Native Hawaiians (MEC-NH) were merged with the 1000 Genomes Project populations. At K = 4 we identified an ancestral component (colored red) that are found largely in Native Hawaiians, presumed to be the Polynesian ancestry.
(TIF)

S9 Fig. Global ancestry proportions estimated from unsupervised ADMIXTURE from K = 4 to K = 8. At each K, ancestry proportions from the replicate (out of five) with highest estimated likelihood output by ADMIXTURE were visualized using Pong. Notably, the inferred proportion of PNS component in Native Hawaiians (red component at K = 4) remains stable across higher K.
(TIF)

S10 Fig. The marginal and cumulative distribution of null test statistics to determine significance threshold for admixture mapping in Native Hawaiians. We used 1,000 runs of genome-wide permutation (top) or 10,000 runs of simulation of test statistics using STEAM (bottom) to determine the distribution of admixture mapping test statistics under the null hypothesis given the correlation structure of estimated local ancestry in MEC Native Hawaiians. The significance threshold was then set as the P-value threshold in which we would obtain a 5% false discovery rate. The threshold was $2.24\text{x}10^{-5}$ using permutation, or $2.28\text{x}10^{-5}$ using STEAM. We thus adopt a threshold of $2.2\text{x}10^{-5}$ for our study (**Methods**).
(TIF)

S1 Table. Details of the association statistics of the covariates and global ancestries of BMI. Model 1 models the non-genetic covariates according to the heuristic described in the **Methods**. The residual from model 1 is then inverse normalized and tested in model 2. Models 1A and 2A repeats the procedure but included quintiles of nSES levels in a mixed effect model (**Methods**); in this case, the R^2 in Model 1A reported include both the fixed and the random effect. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and BMI. See S21 Table for description of these education levels.
(DOCX)

S2 Table. Details of the association statistics of the covariates and global ancestries of WHR. Model 1 models the non-genetic covariates according to the heuristic described in the **Methods**. The residual from model 1 is then inverse normalized and tested in model 2. The top panels were conducted in males only; the bottom in females only. See S21 Table for description of these education and cigarette smoking levels.
(DOCX)

S3 Table. Details of the association statistics of the covariates and global ancestries of fasting glucose. Model 1 models the non-genetic covariates according to the heuristic described

in the **Methods**. The residual from model 1 is then inverse normalized and tested in model 2.
(DOCX)

S4 Table. Details of the association statistics of the covariates and global ancestries of fasting insulin. Model 1 models the non-genetic covariates according to the heuristic described in the **Methods**. The residual from model 1 is then inverse normalized and tested in model 2.
(DOCX)

S5 Table. Details of the association statistics of the covariates and global ancestries of HDL. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. The residual from model 1 is then inverse normalized and tested in model 2. Models 1A and 2A repeats the procedure but included quintiles of nSES levels in a mixed effect model (**Methods**); in this case, the R^2 in Model 1A reported include both the fixed and the random effect.
(DOCX)

S6 Table. Details of the association statistics of the covariates and global ancestries of LDL. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. The residual from model 1 is then inverse normalized and tested in model 2.
(DOCX)

S7 Table. Details of the association statistics of the covariates and global ancestries of TG. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. The residual from model 1 is then inverse normalized and tested in model 2. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and BMI.
(DOCX)

S8 Table. Details of the association statistics of the covariates and global ancestries of total cholesterol. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. The residual from model 1 is then inverse normalized and tested in model 2.
(DOCX)

S9 Table. Details of the association statistics of the covariates and global ancestries of obesity. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. Model 2 then includes global ancestries in addition to the significant covariates. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and obesity. Model 3 included quintiles of nSES levels in a mixed effect model.
(DOCX)

S10 Table. Details of the association statistics of the covariates and global ancestries of Type-2 Diabetes. Model 1 models the non-genetic covariates according to the heuristic described in the **Methods**. Model 2 then includes global ancestries in addition to the significant covariates. Model 3 included quintiles of nSES levels in a mixed effect model. * edu3 was a ternary variable created from the original categorical variable of education status by grouping levels 1 and 2. This was done because there were no significant associations between education levels 1 and 2 with T2D.
(DOCX)

S11 Table. Details of the association statistics of the covariates and global ancestries of heart failure. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. Model 2 then includes global ancestries in addition to the significant covariates. Model 3 included quintiles of nSES levels in a mixed effect model. * edu3 was a ternary variable created from the original categorical variable of education status by grouping levels 1 and 2. This was done because there were no significant associations between education levels 1 and 2 with heart failure.
(DOCX)

S12 Table. Details of the association statistics of the covariates and global ancestries of hyperlipidemia. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. Model 2 then includes global ancestries in addition to the significant covariates. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and hyperlipidemia.
(DOCX)

S13 Table. Details of the association statistics of the covariates and global ancestries of hypertension. Model 1 models the non-genetic covariates according to the heuristic described in the **Methods**. Model 2 then includes global ancestries in addition to the significant covariates.
(DOCX)

S14 Table. Details of the association statistics of the covariates and global ancestries for ischemic heart disease. Model 1 models the non-genetic covariates according to the heuristic described in the Methods. Model 2 then includes global ancestries in addition to the significant covariates. * edu3 was a ternary variable created from the original categorical variable of education status by grouping levels 1 and 2. This was done because there were no significant associations between education levels 1 and 2 with ischemic heart disease.
(DOCX)

S15 Table. Details of the association statistics of the covariates and global ancestries for stroke and transient ischemic attacks. Model 1 models the non-genetic covariates according to the heuristic described in the **Methods**. Model 2 then includes global ancestries in addition to the significant covariates.
(DOCX)

S16 Table. Stratified analysis and interaction between ancestry components and neighborhood SES measures. For traits we analyzed in Table 2 accounting for the impact due to neighborhood SES, we adopted the same model (**S1, S5 and S9–S11 Tables**) but dichotomized the nSES variable, grouping quintiles 1–3 into the "low" nSES group, and quintiles 4–5 into the "high" nSES group, and included the interaction terms between each of the ancestry component and the dichotomized nSES variable. Mixed effect modeling was then performed. The effect on BMI associated with each ancestry in the low and high nSES groups are provided, as well as the P-value for the interaction terms (P_{het}). We observed no significant interaction between ancestry and dichotomized nSES measure.
(DOCX)

S17 Table. Model of association between global ancestry and BMI, including interaction with sex. Model 1 models the non-genetic covariates according to the heuristic described in the Methods, except for sex. The residual from model 1 was then inverse normalized and tested

in model 2, which includes global ancestries, sex, and interactions between global ancestries and sex. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and BMI.
(DOCX)

S18 Table. Stratified analysis of association between global genetic ancestry and BMI among T2D cases and controls. Model testing was performed in the same manner as the global analysis with BMI (**S1 Table**), except for stratifying based on T2D disease status. Model column provided the final model with association coefficients. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and BMI.
(DOCX)

S19 Table. Model of association between global ancestry and BMI, including interaction with type-2 diabetes. Model 1 models the non-genetic covariates according to the heuristic described in the Methods, except for type-2 diabetes status. The residual from model 1 was then inverse normalized and tested in model 2, which includes global ancestries, type-2 diabetes status, and interactions between global ancestries and type-2 diabetes status. * edu4 was a binary variable created from the original categorical variable of education status by grouping levels 1,2,3 and coded 0, while education status level 4 was coded as 1. This was done because there were no significant associations between education levels 1 through 3 and BMI.
(DOCX)

S20 Table. Variants within the admixture signal region that were reported to be associated with the tested or related traits in GWAS catalog. Reported P-value, associated trait, and mapped genes were provided by the GWAS catalog. Allele frequencies were either calculated from the imputed data of the 178 reference MEC Native Hawaiian individuals with estimated PNS ancestry > 90%, or obtained from 1000 Genomes Project. Frequencies were reported with respect to the minor allele in the Native Hawaiians, given in parenthesis next to the Native Hawaiian frequency estimates.
(DOCX)

S21 Table. Allele frequencies across populations for the most strongly associated variant in chr6 for T2D in single variant association test. Allele frequencies were either calculated from the imputed data of the 178 reference MEC Native Hawaiian individuals with estimated PNS ancestry > 90%, or obtained from 1000 Genomes Project (EUR, EAS, and AFR; https://www.internationalgenome.org/1000-genomes-browsers/) or the Genome Asia data (Oceania and Southeast Asia; https://browser.genomeasia100k.org/). Frequencies were reported with respect to the derived allele, given in parenthesis next to the Native Hawaiian frequency estimates.
(DOCX)

S22 Table. Association results to T2D in 2,852 Samoan Replication Cohort. We attempted to replicate the association of rs370140172 and nine other proxies showing the strongest single-variant associations with a cross-sectional population based study of Samoans recruited from Independent Samoa (**Methods**). EAF, effect allele frequency in Samoans. BETA and SE refers to the effect size and standard errors, respectively, from the logistic mixed model association tests in the Samoan cohort. P-val (Samoa) and P-val (MEC-NH) provide the p-value

from the logistic mixed model association tests in the Samoan cohort and MEC Native Hawaiian cohort, respectively.
(DOCX)

S23 Table. Local ancestry inference using RFMix is robust to the choice of recombination map. To evaluate the impact of recombination map on local ancestry inference, we used the 1000 Genomes AMR population. Following the same procedure used for Native Hawaiians, we identified through unsupervised ADMIXTURE analysis 49 Peruvian (PEL) and 3 Mexican (MEX) individuals from 1000 Genomes as having > 80% Native American ancestry. We then inferred local ancestry using RFMix in 71 HapMap3 MEX individuals using the constructed reference panel of 99 CEU, 108 YRI, and 52 NA individuals from 1000 Genomes. We used three recombination map in the local ancestry inference: a HapMap2 pooled recombination map, a mis-specified African-American map, and a constant map that assumes a constant rate of 1cM/Mb across the genome. We compared in pairwise fashion the concordance of inferred ancestry across common variants between runs, and calculated concordance rate as the sum of the diagonal of the contingency table. Across all comparisons, even when using a constant rate map, the concordance rate is extremely high (0.987, 0.981, and 0.981 for the comparisons of default vs. AA map, default to constant rate map, and constant rate to AA map, respectively), suggesting that the choice of recombination map does not strongly impact the local ancestry inference using RFMix.
(DOCX)

S24 Table. Phenotype inclusion and transformation for metabolic and quantitative cardiovascular traits. These traits were studied in PAGE consortium and we thus follow the same criteria and transformation.
(DOCX)

S25 Table. Descriptive summary statistics of the traits and covariates analyzed. Summary statistics reported after exclusion and transformation as described in S20 Table. For biomarkers (glucose, insulin, HDL, LDL, TG, and TC), a subset of participants were invited after cohort entry. Thus there is an age at baseline and an age at blood draw.
(DOCX)

Acknowledgments

We would like to thank all of the Native Hawaiian participants in the Multiethnic Cohort. We also would like to thank University of Hawaiʻi Cancer Center's Native Hawaiian Community Advisory Board for reviewing the study proposal and providing comments to earlier version of this manuscript. Furthermore, we would like to thank the Samoan participants of the study, local village authorities, and the many Samoan and other field workers who contributed to the success of the Samoan OLaGA Study Group. The Samoan OLaGA Study Group includes additional members: Jenna C. Carlson, Jerry Z. Zhang, Nicola L. Hawley, Rajan Deka, Daniel E. Weeks, and Stephen T. McGarvey. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Finally, we thank Gertraud Maskarinec and Hansong Wang for helpful discussions and technical assistance.

Author Contributions

Conceptualization: Charleston W. K. Chiang.

Data curation: Meng Lin, Ryan L. Minster, Bryan L. Dinh, Take Naseri, Muagututiʻa Sefuiva Reupena, Lynne R. Wilkens.

Formal analysis: Hanxiao Sun, Meng Lin, Emily M. Russell, Tsz Fung Chan, Bryan L. Dinh.

Investigation: Charleston W. K. Chiang.

Resources: Take Naseri, Muagututi'a Sefuiva Reupena, Annette Lum-Jones, Iona Cheng, Lynne R. Wilkens, Christopher A. Haiman.

Supervision: Ryan L. Minster, Iona Cheng, Loïc Le Marchand, Christopher A. Haiman, Charleston W. K. Chiang.

Validation: Emily M. Russell.

Visualization: Hanxiao Sun, Meng Lin, Charleston W. K. Chiang.

Writing – original draft: Hanxiao Sun, Meng Lin, Charleston W. K. Chiang.

Writing – review & editing: Hanxiao Sun, Emily M. Russell, Ryan L. Minster, Tsz Fung Chan, Annette Lum-Jones, Iona Cheng, Lynne R. Wilkens, Loïc Le Marchand, Christopher A. Haiman, Charleston W. K. Chiang.

References

1. Hixson L, Hepler BB, Kim MO. The Native Hawaiian and Other Pacific Islander Population: 2010. 2010 Census Briefs. United States Census Bureau; 2012. Available from: https://www.census.gov/prod/cen2010/briefs/c2010br-12.pdf
2. Braden KW, Nigg CR. Modifiable Determinants of Obesity in Native Hawaiian and Pacific Islander Youth. Hawaii J Med Public Health. 2016; 75: 162–71. PMID: 27413626
3. Madan A, Archambeau OG, Milsom VA, Goldman RL, Borckardt JJ, Grubaugh AL, et al. More than black and white: differences in predictors of obesity among Native Hawaiian/Pacific Islanders and European Americans. Obesity (Silver Spring). 2012; 20: 1325–8. https://doi.org/10.1038/oby.2012.15 PMID: 22286530
4. Maskarinec G, Erber E, Grandinetti A, Verheus M, Oum R, Hopping BN, et al. Diabetes incidence based on linkages with health plans: the multiethnic cohort. Diabetes. 2009; 58: 1732–8. https://doi.org/10.2337/db08-1685 PMID: 19258435
5. Pike MC, Kolonel LN, Henderson BE, Wilkens LR, Hankin JH, Feigelson HS, et al. Breast cancer in a multiethnic cohort in Hawaii and Los Angeles: risk factor-adjusted incidence in Japanese equals and in Hawaiians exceeds that in whites. Cancer Epidemiol Biomarkers Prev. 2002; 11: 795–800. PMID: 12223421
6. Singh GK, Lin SC. Dramatic Increases in Obesity and Overweight Prevalence among Asian Subgroups in the United States, 1992–2011. ISRN Prev Med. 2013; 2013: 898691. https://doi.org/10.5402/2013/898691 PMID: 24967142
7. Mau MK, Sinclair K, Saito EP, Baumhofer KN, Kaholokula JK. Cardiometabolic health disparities in native Hawaiians and other Pacific Islanders. Epidemiol Rev. 2009; 31: 113–129. https://doi.org/10.1093/ajerev/mxp004 PMID: 19531765
8. Tung WC, Barnes M. Heart Diseases Among Native Hawaiians and Pacific Islanders. Home Health Care Management and Practice. 2014; 26: 110–113. https://doi.org/10.1177/1084822313516125
9. Grandinetti A, Chen R, Kaholokula JK, Yano K, Rodriguez BL, Chang HK, et al. Relationship of blood pressure with degree of Hawaiian ancestry. Ethn Dis. 2002; 12: 221–8. PMID: 12019931
10. Claw KG, Anderson MZ, Begay RL, Tsosie KS, Fox K, Garrison NA, et al. A framework for enhancing ethical genomic research with Indigenous communities. Nat Commun. 2018; 9: 2957. https://doi.org/10.1038/s41467-018-05188-3 PMID: 30054469
11. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016; 538: 161–164. https://doi.org/10.1038/538161a PMID: 27734877
12. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 2019; 51: 584–591. https://doi.org/10.1038/s41588-019-0379-x PMID: 30926966
13. Diamond JM. Taiwan's gift to the world. Nature. 2000; 403: 709–10. https://doi.org/10.1038/35001685 PMID: 10693781

14. Kim SK, Gignoux CR, Wall JD, Lum-Jones A, Wang H, Haiman CA, et al. Population genetic structure and origins of Native Hawaiians in the multiethnic cohort study. PLoS One. 2012; 7: e47881. https://doi.org/10.1371/journal.pone.0047881 PMID: 23144833

15. Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, et al. Genomic insights into the peopling of the Southwest Pacific. Nature. 2016; 538: 510–513. https://doi.org/10.1038/nature19844 PMID: 27698418

16. Nordyke EC. The Peopling of Hawaii. 2nd ed. University of Hawaii Press; 1989.

17. Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annu Rev Genomics Hum Genet. 2010; 11: 65–89. https://doi.org/10.1146/annurev-genom-082509-141523 PMID: 20594047

18. Mersha TB. Mapping asthma-associated variants in admixed populations. Front Genet. 2015; 6: 292. https://doi.org/10.3389/fgene.2015.00292 PMID: 26483834

19. Shriner D. Overview of Admixture Mapping. Curr Protoc Hum Genet. 2017; 94: 1.23.1–1.23.8. https://doi.org/10.1002/cphg.44 PMID: 28696560

20. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. 2000; 151: 346–57. https://doi.org/10.1093/oxfordjournals.aje.a010213 PMID: 10695593

21. Wojcik GL, Fuchsberger C, Taliun D, Welch R, Martin AR, Shringarpure S, et al. Imputation-Aware Tag SNP Selection To Improve Power for Large-Scale, Multi-ethnic Association Studies. G3 (Bethesda). 2018. https://doi.org/10.1534/g3.118.200502 PMID: 30131328

22. Conroy SM, Shariff-Marco S, Yang J, Hertz A, Cockburn M, Shvetsov YB, et al. Characterizing the neighborhood obesogenic environment in the Multiethnic Cohort: a multi-level infrastructure for cancer health disparities research. Cancer Causes Control. 2018; 29: 167–183. https://doi.org/10.1007/s10552-017-0980-1 PMID: 29222610

23. Lim U, Monroe KR, Buchthal S, Fan B, Cheng I, Kristal BS, et al. Propensity for Intra-abdominal and Hepatic Adiposity Varies Among Ethnic Groups. Gastroenterology. 2019; 156: 966–975.e10. https://doi.org/10.1053/j.gastro.2018.11.021 PMID: 30445012

24. GenomeAsia100K Consortium. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature. 2019; 576: 106–111. https://doi.org/10.1038/s41586-019-1793-z PMID: 31802016

25. Lin M, Caberto C, Wan P, Li Y, Lum-Jones A, Tiirikainen M, et al. Population specific reference panels are crucial for the genetic analyses of Native Hawai'ians: an example of the CREBRF locus. bioRxiv. 2019; 789073. https://doi.org/10.1101/789073

26. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. Cancer Causes Control. 2001; 12: 703–711. https://doi.org/10.1023/a:1011240019516 PMID: 11562110

27. Von Behren J, Abrahão R, Goldberg D, Gomez SL, Setiawan VW, Cheng I. The influence of neighborhood socioeconomic status and ethnic enclave on endometrial cancer mortality among Hispanics and Asian Americans/Pacific Islanders in California. Cancer Causes Control. 2018; 29: 875–881. https://doi.org/10.1007/s10552-018-1063-7 PMID: 30056614

28. DeRouen MC, McKinley M, Shah SA, Borno HT, Aoki R, Lichtensztajn DY, et al. Testicular cancer in Hispanics: incidence of subtypes over time according to neighborhood sociodemographic factors in California. Cancer Causes Control. 2020; 31: 713–721. https://doi.org/10.1007/s10552-020-01311-2 PMID: 32440828

29. Swinburn BA, Ley SJ, Carmichael HE, Plank LD. Body size and composition in Polynesians. Int J Obes Relat Metab Disord. 1999; 23: 1178–1183. https://doi.org/10.1038/sj.ijo.0801053 PMID: 10578208

30. Gastaldelli A, Miyazaki Y, Pettiti M, Matsuda M, Mahankali S, Santini E, et al. Metabolic effects of visceral fat accumulation in type 2 diabetes. J Clin Endocrinol Metab. 2002; 87: 5098–5103. https://doi.org/10.1210/jc.2002-020696 PMID: 12414878

31. Neeland IJ, Turer AT, Ayers CR, Powell-Wiley TM, Vega GL, Farzaneh-Far R, et al. Dysfunctional adiposity and the risk of prediabetes and type 2 diabetes in obese adults. JAMA. 2012; 308: 1150–1159. https://doi.org/10.1001/2012.jama.11132 PMID: 22990274

32. Grinde KE, Brown LA, Reiner AP, Thornton TA, Browning SR. Genome-wide Significance Thresholds for Admixture Mapping Studies. Am J Hum Genet. 2019; 104: 454–465. https://doi.org/10.1016/j.ajhg.2019.01.008 PMID: 30773276

33. Greenwald WW, Chiou J, Yan J, Qiu Y, Dai N, Wang A, et al. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. Nat Commun. 2019; 10: 2078. https://doi.org/10.1038/s41467-019-09975-4 PMID: 31064983

34. Abd El-Aziz MM, Barragan I, O'Driscoll CA, Goodstadt L, Prigmore E, Borrego S, et al. EYS, encoding an ortholog of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa. Nat Genet. 2008; 40: 1285–1287. https://doi.org/10.1038/ng.241 PMID: 18836446

35. Littink KW, van den Born LI, Koenekoop RK, Collin RWJ, Zonneveld MN, Blokland EAW, et al. Mutations in the EYS gene account for approximately 5% of autosomal recessive retinitis pigmentosa and cause a fairly homogeneous phenotype. Ophthalmology. 2010; 117: 2026–2033, 2033.e1–7. https://doi.org/10.1016/j.ophtha.2010.01.040 PMID: 20537394

36. Scott RA, Scott LJ, Mägi R, Marullo L, Gaulton KJ, Kaakinen M, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes. 2017; 66: 2888–2902. https://doi.org/10.2337/db16-1253 PMID: 28566273

37. Akiyama M, Okada Y, Kanai M, Takahashi A, Momozawa Y, Ikeda M, et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat Genet. 2017; 49: 1458–1467. https://doi.org/10.1038/ng.3951 PMID: 28892062

38. Maskarinec G, Morimoto Y, Jacobs S, Grandinetti A, Mau MK, Kolonel LN. Ethnic admixture affects diabetes risk in native Hawaiians: the Multiethnic Cohort. Eur J Clin Nutr. 2016; 70: 1022–1027. https://doi.org/10.1038/ejcn.2016.32 PMID: 27026423

39. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019; 570: 514–518. https://doi.org/10.1038/s41586-019-1310-4 PMID: 31217584

40. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet. 2016; 98: 127–48. https://doi.org/10.1016/j.ajhg.2015.11.022 PMID: 26748516

41. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013; 93: 278–88. https://doi.org/10.1016/j.ajhg.2013.06.020 PMID: 23910464

42. Wegmann D, Kessner DE, Veeramah KR, Mathias RA, Nicolae DL, Yanek LR, et al. Recombination rates in admixed individuals identified by ancestry-based inference. Nat Genet. 2011; 43: 847–853. https://doi.org/10.1038/ng.894 PMID: 21775992

43. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, et al. The landscape of recombination in African Americans. Nature. 2011; 476: 170–175. https://doi.org/10.1038/nature10336 PMID: 21775986

44. Wang H, Haiman CA, Kolonel LN, Henderson BE, Wilkens LR, Le Marchand L, et al. Self-reported ethnicity, genetic structure and the impact of population stratification in a multiethnic study. Hum Genet. 2010; 128: 165–77. https://doi.org/10.1007/s00439-010-0841-4 PMID: 20499252

45. Minster RL, Hawley NL, Su CT, Sun G, Kershaw EE, Cheng H, et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. Nat Genet. 2016; 48: 1049–1054. https://doi.org/10.1038/ng.3620 PMID: 27455349

46. Setiawan VW, Virnig BA, Porcel J, Henderson BE, Le Marchand L, Wilkens LR, et al. Linking data from the Multiethnic Cohort Study to Medicare data: linkage results and application to chronic disease research. Am J Epidemiol. 2015; 181: 917–919. https://doi.org/10.1093/aje/kwv055 PMID: 25841869

47. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010; 26: 2867–73. https://doi.org/10.1093/bioinformatics/btq559 PMID: 20926424

48. Pino-Yanes M, Gignoux CR, Galanter JM, Levin AM, Campbell CD, Eng C, et al. Genome-wide association study and admixture mapping reveal new loci associated with total IgE levels in Latinos. J Allergy Clin Immunol. 2015; 135: 1502–1510. https://doi.org/10.1016/j.jaci.2014.10.033 PMID: 25488688

49. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019; 47: D1005–D1012. https://doi.org/10.1093/nar/gky1120 PMID: 30445434

50. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010; 42: 348–54. https://doi.org/10.1038/ng.548 PMID: 20208533

51. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol. 2015; 39: 276–293. https://doi.org/10.1002/gepi.21896 PMID: 25810074

52. Hawley NL, Minster RL, Weeks DE, Viali S, Reupena MS, Sun G, et al. Prevalence of adiposity and associated cardiometabolic risk factors in the Samoan genome-wide association study. Am J Hum Biol. 2014; 26: 491–501. https://doi.org/10.1002/ajhb.22553 PMID: 24799123

53. Ziyatdinov A, Vázquez-Santiago M, Brunel H, Martinez-Perez A, Aschard H, Soria JM. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics. 2018; 19: 68. https://doi.org/10.1186/s12859-018-2057-x PMID: 29486711

54. Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. On detecting incomplete soft or hard selective sweeps using haplotype structure. Mol Biol Evol. 2014; 31: 1275–1291. https://doi.org/10.1093/molbev/msu077 PMID: 24554778

55. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol Biol Evol. 2014; 31: 2824–2827. https://doi.org/10.1093/molbev/msu211 PMID: 25015648