Recombination map tailored to Native Hawaiians reduces false positive findings in genomic scans of positive selection

Bryan L. Dinh[1,2], Echo Tang[1], Kekoa Taparra[3], Nathan Nakatsuka[4], Fei Chen[2], and Charleston W. K. Chiang[1,2]

[1] Department of Quantitative and Computational Biology, University of Southern California

[2] Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California

[3] Department of Radiation Oncology, Stanford University, Palo Alto, California

[4] New York Genome Center, New York, New York

Abstract

Recombination events establish the patterns of haplotypic structure in a population and estimates of recombination rates are used in several downstream population and statistical genetic analyses. However, the recombination landscape of populations not represented in the 1000 Genomes Project (1KGP) is poorly understood. Using suboptimal maps from distantly related populations reduces the efficacy of extending genomic analyses to underrepresented populations, including Native Hawaiians. To overcome this challenge, we constructed linkage disequilibrium (LD)- and Identity-by-Descent (IBD)-based recombination maps specifically for Native Hawaiians. We created and characterized two recombination maps from our study sample: one based on a random subset of our sample reflecting the current admixed state of Native Hawaiians (NH map), and one based on a subset of individuals with enriched Polynesian ancestries (PNS map) with potential to be used for less admixed Polynesian populations such as the Samoans. We found the recombination landscape to be less correlated with those from other 1KGP populations (e.g. Spearman's rho = 0.79 between PNS and CEU (Utah residents with Northern and Western European ancestry) compared to 0.92 between YRI (Yoruba in Ibadan, Nigeria) and CEU at 50 kb resolution. PNS shared the fewest recombination hotspots with other populations from 1KGP (e.g. 8% of hotspots shared between PNS and CEU compared to 27% of hotspots shared between YRI and CEU). We found that downstream analysis in the Native Hawaiian population, such as local ancestry inference, imputation, and IBD segment and relatedness detections, would achieve similar efficacy when using the NH map compared to an omnibus map. However, for genome scans of adaptive loci using integrated haplotype scores,

32 we found several loci with apparent genome-wide significant signals (|Z-score| > 4) in Native
33 Hawaiians that would not have been significant when analyzed using NH-specific maps.
34 Therefore, population-specific recombination maps improve the robustness of haplotype-based
35 statistics and help us better characterize the evolutionary history that may underlie Native
36 Hawaiian-specific health conditions that persist today.

Introduction

Knowledge of the recombination landscape across the genome informs patterns of haplotypic structure in a population and is used in several population and statistical genetic analyses such as imputation, local ancestry inference (LAI), identity-by-descent (IBD) inference, and genomic scan of adaptive loci, among others. Fine-scale differences in recombination landscapes are known to exist between populations (1–5). For example, recombination maps for HapMap populations CEU (Utah residents of predominantly Northern and Western European ancestries) and YRI (Yoruba in Ibadan, Nigeria) correlate well at large resolutions but have poorer correlation at finer scales (3). Moreover, the deCODE map, based on individuals with European-ancestries from Iceland, correlates better with CEU than YRI (3). Any differences in recombination landscapes between populations can be due to both the divergent demographic histories as well as genetic differences in usage of recombination hotspots (1). However, we have limited understanding of the recombination landscape for populations not represented by the 1000 Genomes Project (1KGP) (6), although efforts in developing these maps are starting for diverse populations such as the Nama people of southern Africa (5).

Given the prevalent use of recombination maps in genetic analysis and the fine-scale differences between populations, it may be suboptimal to use a recombination map from a distantly related population. This may further exacerbate the known disparity due to an underdevelopment of genomic resources for analysis of underrepresented populations. However, the benefit of a population specific recombination map on downstream analyses has

58 not been extensively explored, and studies have reached different qualitative conclusions. For
59 an isolated, homogenous population like the Finns (4), a population-specific map did not show a
60 large impact on downstream analyses such as phasing and imputation, although this could be
61 due to general haplotypic similarity between Finns and the abundance of European-ancestry
62 samples represented in 1000 Genomes or that information from recombination plays only a
63 secondary role in the specific analyses evaluated. On the other hand, population-specific
64 recombination maps have also been shown to impact haplotype-based scans of positive
65 selection in the Nama (5), a population not adequately represented by available maps
66 emphasizing West African ancestry.
67
68 In this work, we focused on characterizing the recombination landscape for the Native
69 Hawaiians, using genome-wide genotyped individuals from the Multiethnic Cohort (7). Native
70 Hawaiians are an admixed, indigenous, and underrepresented population that, with a current
71 population of 680,353 individuals, account for only 0.2% of the U.S. population (8). The lack of
72 resources to advance genomic research among Native Hawaiians has previously been detailed
73 (9,10). We thus also explored the impact of a population-specific map on multiple downstream
74 analyses for the Native Hawaiians. The lack of available whole genome sequencing data and the
75 relatively small cohort size limited the applicable methods that could be used to accurately
76 infer a recombination for this population. LDhat has been shown to infer accurate
77 recombination maps using array data with good concordance to maps inferred through other
78 methods (11,12). Because of European colonization and subsequent waves of immigrations to
79 the Hawaiian islands, we modeled the Native Hawaiians to include African-, East Asian-, and

80 European-related ancestries in addition to Polynesian ancestries following previous work of this
81 population (9). As a result, we created and characterized two recombination maps from our
82 study sample: one based on a random subset of our sample reflecting the current admixed
83 state of Native Hawaiians, and one based on a subset of individuals with enriched Polynesian
84 ancestries. The latter map was constructed so to gain insights on the recombination landscape
85 of ancestral Polynesians and to potentially provide a viable map for relatively unadmixed
86 Polynesian populations such as the Samoans. We then evaluated the impact of a Native
87 Hawaiian-specific map for downstream analyses such as LAI, imputation, IBD segment
88 inference, and genome-wide scans of adaptation using the integrated haplotype score.

Results

Benchmarking the recombination map inference pipeline

We implemented a pipeline to generate recombination maps using LDhat (**Materials and Methods**). Notably, LDhat infers the relative, population-level, recombination rate, ρ. However, following previous protocols, we inferred and regressed out the contribution of the demographic history as measured by effective population size, N_e (**Materials and Methods**). We benchmarked our pipeline by applying it to OMNI array genotypes for 96 randomly selected CEU individuals from 1KGP and comparing the resulting map from our pipeline to the publicly available map originally also inferred from the OMNI array data. Overall, we observed high correlations (r = 0.93 to 0.96) across all scales of the map, ranging from 50 kb to 5 Mb (**Table 1**). The correlation did lower at the finest scale of 10 kb (r = 0.86), likely due to noise in LD estimates at fine scales given finite sample size (13). We also extracted the set of OMNI array SNPs from the public release of 1KGP phase 3 sequencing data for the CEU population to infer the recombination map, which again showed high correlation across all scales when compared to that inferred from the OMNI array genotype using the same in-house pipeline (r = 0.91 – 0.98; **Table 1**). Therefore, we conclude that our implemented pipeline near-faithfully recreated the previously published map, and that, given the same SNP content, differences in the distribution of any genotyping errors attributed to data generation platform (OMNI array vs. low pass sequencing) did not substantially impact the recombination map inference.

Table 1: Benchmarking LDhat inference pipeline

Inferred Map Data Source			Comparison Data Source	Resolution				
Population	SNP Set	Data Source		10kb	50kb	100kb	1Mb	5Mb
CEU	OMNI	Array	Published	0.86	0.93	0.94	0.95	0.96
CEU	OMNI	WGS	OMNI array data, in-house	0.91	0.95	0.95	0.96	0.98
CEU	MEGA	WGS	OMNI SNPs from WGS data, in-house	0.78	0.91	0.93	0.94	0.97
YRI	MEGA	WGS	Published	0.82	0.92	0.94	0.95	0.97

Each map was inferred from 96 randomly selected individuals from the corresponding 1KGP population via our LDhat pipeline. The inferred map would either be based on OMNI array genotypes, OMNI array SNPs extracted from WGS genotypes, or MEGA array SNPs extracted from WGS genotypes. The inferred maps were compared to the corresponding 1KGP recombination map for the same population or another inferred map using our in-house pipeline. We measured concordance between maps by the Pearson correlation coefficient. The first comparison established that our implemented pipeline faithfully recreated the published recombination map for CEU using the same input dataset. The second comparison established that the inferred map is robust to any differences in genotype quality between genotyping or sequencing. The third comparison established that maps based on the set of OMNI array SNPs or MEGA array SNPs are highly concordant, at least for 50kb scale or higher. The last comparison in YRI suggested that maps inferred using MEGA array genotypes can be compared directly to published 1KGP maps.

Because the Native Hawaiian genetic data from the Multiethnic Cohort is only available on the MEGA array, we thus also benchmarked the impact of map inference using a different set of SNPs. Again, using the CEU 1KGP population as the example, we extracted the genotypes found on the MEGA array to infer a recombination map. The resulting map is highly correlated with the version generated based on extracted OMNI array SNPs, at least for 50 kb scale or higher ($r > 0.91$; **Table 1**). Taken together, our LDhat inference pipeline, when applied to SNPs found on the MEGA array for the CEU population would produce highly concordant recombination maps at 50 kb scale or higher compared to the published map. Therefore, the recombination landscape for Native Hawaiians that we inferred here can be compared to those published for other 1KGP populations. Nevertheless, for between-population comparisons of the recombination landscape we compared the inferred NH maps to those based on the MEGA array SNP content extracted from 1KGP sequence data.

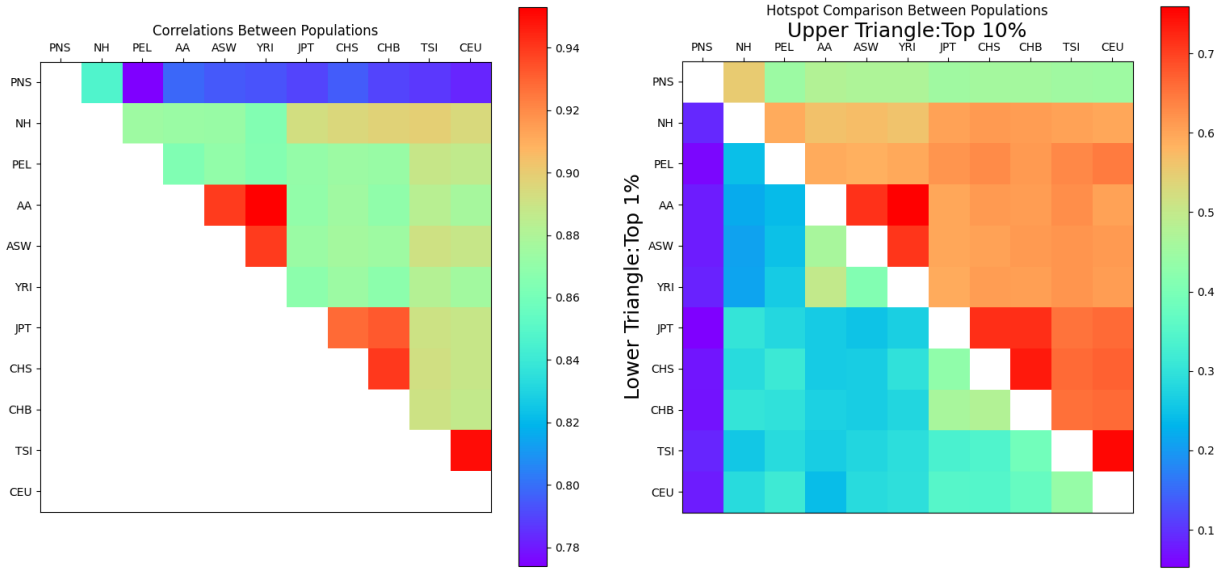
Contrasting the Native Hawaiian recombination landscape

We proceeded to infer a recombination map for 96 randomly selected admixed Native Hawaiians from our study sample (NH map) to represent the current Native Hawaiian population, and a map for 96 individuals from our study sample previously estimated to be most enriched with the Polynesian ancestries found predominantly in Native Hawaiians (PNS map; **Materials and Methods**). The PNS map was constructed to provide insights on the recombination landscape of the ancient Polynesians and how it may differ with other continental populations around the world. However, we caution that the Polynesian ancestries found predominantly in Native Hawaiians may or may not be identical with that found in other Polynesian populations such as the Samoans. We then compared the inferred Native Hawaiian recombination landscape with that from other populations that were similarly constructed using MEGA array genotypes extracted from the 1KGP phase 3 datasets.

We found that populations with recent shared history displayed higher levels of correlation, as expected (**Figure 1a**). This is notable for populations representing each of the ancestral components found: African (MEC AA, ASW, and YRI), East Asian (CHB, CHS, and JPT), and European (CEU, TSI). At the 50 kb scale, the recombination landscape as indicated by the NH map showed intermediate correlations with the non-Native Hawaiian populations (r = 0.865 - 0.899), tending to correlate best with populations representing East Asian (r = 0.892 – 0.898) and European ancestry (r = 0.895 - 0.899), likely due to substantial admixture from these two ancestral components. In contrast, the landscape from PNS showed the lowest correlations

156 with almost all populations compared (r = 0.774 - 0.847). PNS correlated most strongly with NH
157 (r = 0.847) and most weakly with Peruvian in Lima, Peru (PEL; r = 0.774). Overall, PNS correlated
158 more poorly with other populations than PEL, a population with a known history of isolation.
159 Taken together, our results are consistent with PNS representing a previously unrepresented
160 component of ancestry by the 1000 Genomes and may suggest a stronger effect of isolation or
161 enhanced genetic drift in the past compared to PEL (although we also note that we did not
162 prune PEL to enrich for Indigenous American ancestry due to small sample size).
163
164 These trends were also observed when we compared recombination hotspot sharing between
165 populations (**Figure 1b**). Following methodology of previous studies (3), we divided the genome
166 into 50 kb windows and defined hotspots for each population as the windows in the top 1
167 percentile or 10 percentile with the highest average recombination rate. For each pair of
168 populations, we then compared the shared proportion of hotspots. The relationships at both
169 hotspot thresholds were similar to the ones observed for correlations of the genome-wide
170 recombination landscape (**Figure 1a**): populations expected to represent recent shared history
171 had the highest sharing at both thresholds. In contrast, PNS had lower hotspot sharing relative
172 to other populations except for NH, with the lowest sharing with PEL. Compared to PNS, the NH
173 map has better relative sharing with the other 1KGP populations, including slightly higher
174 sharing with European and East Asian populations. PNS's low correlation genome-wide in
175 recombination rate variations and hotspot sharing with other populations is thus indicative of a
176 unique recombination landscape.
177

178 **Figure 1: Correlation of recombination landscape and hotspots between populations.**



179
180 Correlations between populations at 50 kb scale (**A**) and sharing of recombination hotspots defined as the
181 top 1% (lower left) or 10% (upper right) windows 50 kb in size by inferred recombination rate (**B**). Both the
182 correlations and proportion of sharing were calculated by excluding genomic regions that fall within the
183 analysis masks (**Materials and Methods**). All landscape and hotspot comparisons here were based on
184 recombination maps inferred in-house using MEGA array SNPs from genotyping (PNS, NH, AA) or extracted
185 from WGS data (1KGP populations with Indigenous American- [PEL], African- [ASW, YRI], East Asian- [JPT,
186 CHS, CHB], and European-ancestries [TSI, CEU]). AA is African American data from MEC.
187

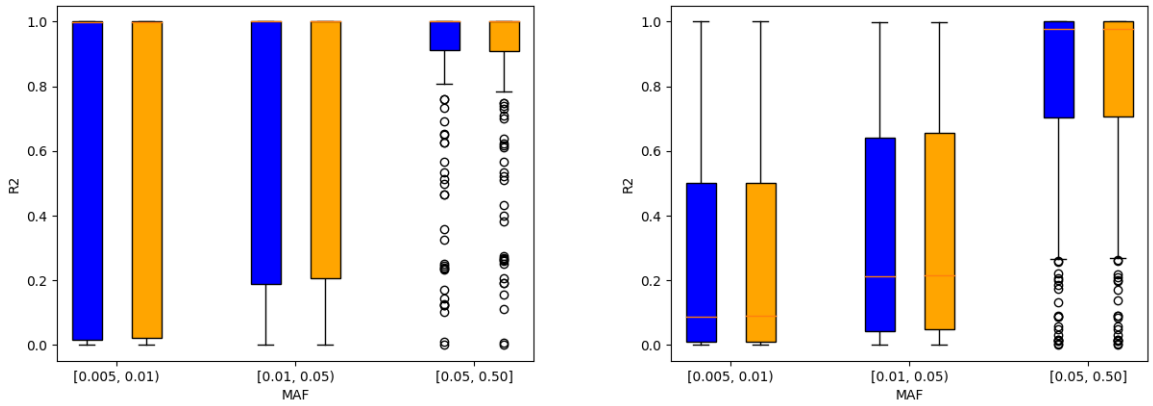
188 ## Impact of the recombination map on imputation

189 We evaluated the impact of a population-specific map for Native Hawaiians on downstream
190 statistical and population genetic applications, specifically focusing on the impact of the
191 recombination map on genotype imputation, local ancestry inference, IBD and relatedness
192 inference, and genomic scans for positive selection.

193

194 Imputation is a common analysis that uses haplotype sharing to infer genotypes at variants not
195 directly observed in a given dataset. The recombination map is needed for phasing (a pre-
196 requisite for imputation) and for imputation itself as it provides the prior that the ancestral

197 haplotype may have switched from one haplotype to another at a given genomic location. We
198 thus evaluated the impact of using a potentially mis-specified omnibus map (*i.e.* the default
199 recombination map released by the commonly used phasing software, Eagle (14,15)) in phasing
200 and imputation, compared to the population-specific map constructed in the present study.
201
202 We imputed 453 and 154 Native Hawaiian individuals genotyped on the Illumina Human660W
203 (NHBC) and MEGA arrays, respectively, against the 1KGP+HGDP reference panel released by
204 gnomAD (16). Compared to the targeted exon sequencing data we have on overlapping
205 individuals, we were able to evaluate imputation accuracy in the 658 and 482 SNPs for NHBC
206 and MEGA cohorts, respectively. Across three minor allele frequency (MAF) bins, 0.5 – 1%, 1 –
207 5%, and 5 – 50%, we observed no notable difference in imputation accuracy when imputation
208 was performed using the NH map compared to the omnibus map (**Figure 2**, **Supplemental**
209 **Table 1**).
210
211 **Figure 2: evaluation of the impact of recombination maps on imputation accuracy**



212
213 Comparison of imputation accuracy using the NH map (blue) and the omnibus map (orange). MAF for each
214 SNP was computed using the sequencing data. **Left**: 154 individuals on the MEGA array were compared on
215 482 available SNPs. Restricting to MAF ≥0.5% resulted in bins with 134, 112, and 168 SNPs, respectively.

Right: 453 individuals on the Human660W array were compared on 658 overlapping SNPs. More SNPs are available for comparison here due to larger number of individuals available for analysis. Restricting to MAF ≥0.5% resulted in bins with 64, 94, and 165 SNPs, respectively.

Impact of recombination map on local ancestry inference

We inferred the local ancestries across 3,665 Native Hawaiian individuals genotyped on the MEGA array using RFMIX (17), with either the NH map or the omnibus map that provided information of recombination rates genome-wide. Overall, there is a large agreement of local ancestries inferred based on the NH map and those based on the omnibus map (concordance = 97.2%, **Table 2**), particularly when phasing was performed only once before local ancestry inference with two separate maps. If we separately phased and ran RFMIX with different maps, the resulting concordance lowered to 95.8% (**Table 2**), suggesting that the choice of the recombination map could make a difference in the LAI calls, though perhaps not strongly. These results are also consistent with previous findings that even a naïve uniform recombination map would produce <10% of discordance in inferred ancestry calls (18).

Table 2: Concordance of inferred local ancestry calls based on different recombination maps

LAI with different maps (data is pre-phased using NH map)

Omnibus map \ NH map	PNS	AFR	EAS	EUR
PNS	49.21%	0.01%	0.92%	0.30%
AFR	0.01%	1.13%	0.00%	0.03%
EAS	1.00%	0.00%	19.21%	0.07%
EUR	0.35%	0.03%	0.07%	27.64%

Phasing and LAI with different maps

	NH map			
	PNS	AFR	EAS	EUR

Omnibus map	PNS	48.61%	0.02%	1.37%	0.48%
	AFR	0.02%	1.10%	0.01%	0.04%
	EAS	1.46%	0.01%	18.69%	0.13%
	EUR	0.49%	0.04%	0.13%	27.39%

The Native Hawaiian population is modeled as a four-way admixture, with ancestry components from Polynesia (PNS), Africa (AFR), East Asia (EAS), and Europe (EUR).

Impact of recombination map on IBD segment calling and relatedness inference

Our initial IBD analysis identified regions of the genome where we observed extreme pileups of IBD segments that are spurious (>90% of individuals would share IBD segments locally). We thus created analysis masks to remove these spurious regions driven by insufficient SNP coverage (see **Materials and Methods**, **Supplemental Figure 1**). When comparing the distribution of IBD segments using masked versions of both maps, we observed little difference (**Supplemental Figure 1**). Given that the inferred recombination rates in regions of sparse SNP density could be unreliable and confound downstream analysis such as IBD segment calling, we also released a set of masked regions (**Supplemental Table 2**) for users to apply when using these recombination maps.

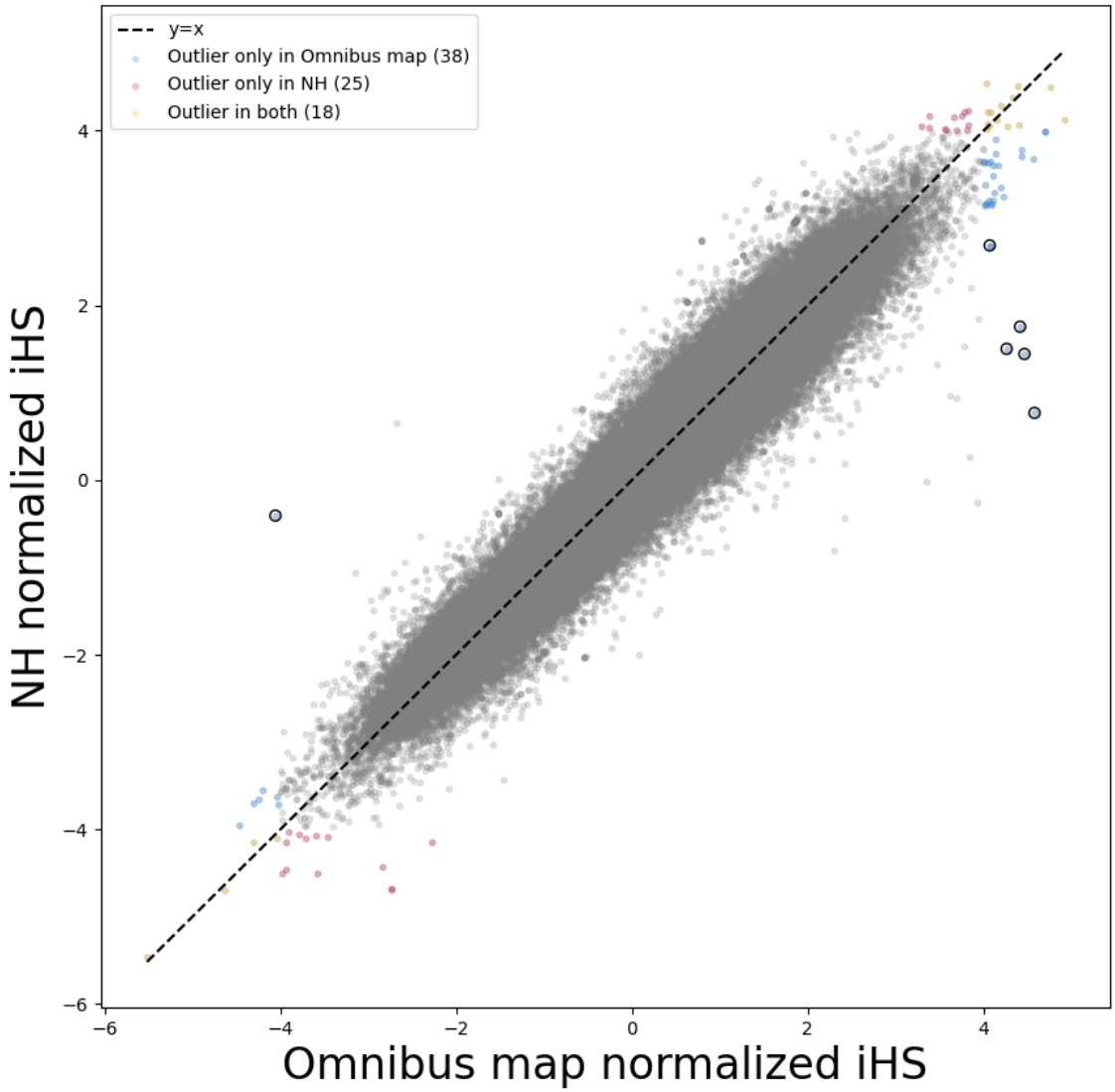
We used the kinship coefficient (φ) based on the pairwise proportion of IBD sharing to compare differences in relatedness inference due to the recombination map (**Materials and Methods**). Of the 218,132 pairs of individuals inferred by KING to be 3rd-degree or closer relatives, 200,617 were found in both the omnibus and NH maps to compare. The estimated kinship coefficient

255 from each map was calculated independently and the maps were found to be highly
256 concordant with a Pearson correlation coefficient of 0.995.
257

258 Impact of recombination map on genomic scans for positive selection
259 using integrated haplotype score

260 Lastly, we evaluated the impact of recombination maps on the robustness of a haplotype-based
261 metric of positive selection, the integrated haplotype score (iHS) (19). We considered loci with
262 normalized |iHS| > 4 as candidate loci under positive selection. While there are candidate loci
263 that are exclusive to either map or are found in both maps, many of these points were close to
264 the identity line with similar iHS values (**Figure 3**). However, we observed multiple SNPs
265 showing extreme iHS values when using the omnibus map but show more attenuated signals
266 when using the NH map. Specifically, we identified 6 SNPs across 4 loci (nearby genes: DISP1,
267 TLCD1, FAM222B, and WNT7B) that showed |iHS| values > 4 when using the omnibus map but
268 normalized |iHS| values between 0 to 2 when using the NH map (**Supplemental Figures 2-5**).
269 Moreover, there are additional loci that would have attenuated evidence of selection in
270 analysis using the NH map (|iHS|< 2) but fall short of our conservative cutoff of 4 standard
271 deviations when using the omnibus map. Therefore, using the population-specific
272 recombination map in haplotype-based selection scans may protect against spurious detections
273 of loci under positive selection.
274

275 **Figure 3: Genomic scan for positive selection using haplotype-based statistics is sensitive to**
276 **choice of recombination map**



Genomic scan of adaptation was measured using iHS as a metric, as implemented in selscan. We estimated iHS for 486,600 SNPs in 150 Native Hawaiian individuals from the MEC genotyped on the MEGA array using either the omnibus map (x-axis) or the NH LDhat (y-axis) map. After normalization by selscan, a score of |z| > 4 is considered genome-wide significant. We identified six SNPs across four loci that would have significant evidence of adaptation using iHS based on the omnibus map, when all of the statistical evidence would be much more attenuated if a population-specific map was used.

A recombination map based on IBD segments

In addition to LDhat, an alternative approach to infer genome-wide recombination maps based on array data is to utilize information from IBD segments. As larger sample size is needed to have sufficient number of IBD segments for inference, we inferred a recombination map using 3,937 individuals using IBDrecomb (11), which would be more akin to the NH map constructed

290 using LDhat. The resulting IBD-based map showed large concordance with the NH map
291 (**Supplemental Table 3**), although the correlation may be lower compared to that reported
292 previously for African Americans (11). For instance, the correlation between an IBD-based map
293 of an African American cohort and an LDhat map from ASW was previously reported to be as
294 high as 0.95 at the 500kb scale; we observed a correlation of 0.90, 0.90, and 0.92 at the 500 kb,
295 1 Mb, and 5 Mb scales (**Supplemental Table 3**). This may be due to the differences in the
296 precision in IBD segment detection using array versus sequencing data (20,21), which may
297 percolate to the precision in the constructed recombination maps, or it may be due to
298 differences in ancestry composition between our African American cohort (sampled largely in
299 Southern California) with that tested previously (11). In comparison, the correlation between
300 IBD-based map and LDhat map from NH are between 0.85 to 0.89 across the same scales
301 (**Supplemental Table 3**). Nevertheless, we found that compared to the omnibus map, the IBD-
302 based map also did not substantially impact the downstream inference for imputation
303 (**Supplemental Figure 6**), LAI (**Supplemental Table 4**; 94.9% concordance, comparing IBD-based
304 to LDhat-based map), and kinship coefficient estimates (Pearson correlation with NH map:
305 0.9945). We also found that genomic scan for positive selection using iHS would be more robust
306 if we were to use an IBD-based recombination map for NH as well (**Supplemental Figure 7**),
307 confirming our previous observation that a population-specific recombination map would be
308 important for performing and interpreting haplotype-based analyses.

Discussion

In the present study we implemented a pipeline to infer recombination maps for populations using LDhat. By utilizing both array and sequence data from 1KGP populations and comparing to the corresponding publicly available recombination maps, we demonstrated that our implemented pipeline could faithfully produce accurate and precise recombination maps. We also showed that recombination maps based on SNP content on the MEGA array have high concordance with published 1KGP maps. These results suggest that our constructed maps for the Native Hawaiian population based on the MEGA array data would be comparable to those already available for 1KGP populations based on the OMNI array.

To characterize the recombination landscape of Native Hawaiians and to generate a commonly used genomic resource for future genetic studies in Polynesian-ancestry individuals, we created two recombination maps, the PNS map and the NH map, from Native Hawaiian individuals estimated to have the highest proportion of Polynesian ancestry and individuals randomly sampled from the cohort, respectively. The former map was constructed to provide insights on the recombination landscape in the ancestral Polynesians prior to European colonization and subsequent waves of immigrations. However, we note that Polynesian ancestries are complex. This component of ancestry is one that is found prevalently in the MEC Native Hawaiian cohort, but could represent a mixture of Polynesian, Austronesian, and other ancestries that are unique to our sample. Therefore, a future evaluation of the applicability of the PNS map to other Polynesian populations across the Pacific would be of interest. The admixture history specific to

330 the Native Hawaiians also contributed to the health disparity experienced by the population
331 today, and the genetic risks to diseases must be evaluated considering this history (18,22). As
332 such, we also constructed the latter map, to better reflect the genomic composition of the
333 Native Hawaiians today. The NH map would be more suited for genetic epidemiologic studies
334 that use a population-based sample and is the primary map that we evaluated throughout this
335 study.
336
337 Overall, we found that the recombination landscape from the PNS map stood out to be the
338 most unique compared to other 1KGP and MEC populations based on the correlation of
339 recombination rate variations across the genome and proportion of hotspot sharing. Notably,
340 the pairwise sharing between YRI and CEU, two populations known to have divergent
341 population histories, have higher correlation with one another (r = 0.876) than either
342 population with PNS (0.793 and 0.783 for YRI and CEU, respectively). The landscape based on
343 the NH map, in contrast, showed higher correlations with the maps from other populations,
344 with slightly higher correlation with the maps inferred from European and East Asian
345 populations. This is likely due to the recent admixture from other continental populations seen
346 within the NH cohort today (18). Lastly, the PNS map showed the lowest correlation with PEL,
347 and both PNS and PEL maps showed relatively low correlations with maps from other
348 populations examined here. The Peruvian from Lima, Peru, is known for its history of isolation
349 relative to other 1KGP populations (23). The observation here thus suggests that long-term
350 isolation and lower effective population sizes are driving the LD pattern and shaping the

351 recombination landscape. The low correlation between PNS and PEL, and between the two with
352 other populations, are potentially reflecting their respective independent isolation histories.
353
354 We evaluated the impact of a population-specific map on downstream population and
355 statistical genetic analyses. In general, we found little difference due to the choice of the
356 recombination maps, particularly for imputation accuracy, local ancestry inference, and IBD
357 segment detections and genetic relatedness inference. In local ancestry inference, we did
358 observe an increased discordance between the two maps, if the input data were also phased
359 separately using different maps (**Table 2**), suggesting that the choice of recombination maps
360 would have a small but tangible impact on these downstream analyses. However, in each of
361 these applications, the recombination maps are used mostly as priors while the actual inference
362 is ultimately driven by the genetic data. Therefore, even though there are fine-scale differences
363 in the recombination landscape between Native Hawaiians and other 1KG populations, the
364 accuracy of these downstream applications will not be substantially impacted by the choice of
365 the map used.
366
367 Our findings suggest that the choice of the recombination map could make qualitative
368 differences in population genetic analyses such as genomic scans for positive selection using
369 the integrated haplotype score (iHS) because these statistics would incorporate recombination
370 information in their calculation, representative of haplotype length surrounding a SNP. We
371 found that six SNPs across 4 loci appeared to represent outliers in an analysis using the omnibus
372 map (using a threshold of |iHS| > 4 after normalization) but have attenuated signals using the

373 NH map (|iHS| < 2 for five of the six SNPs). The signals in the four loci are driven by the genes
374 DISP1, TLCD1, FAM222B, and WNT7B. We examined each locus manually and did not find clear
375 evidence for them to be under selection. DISP1 impacts signaling and transport for cellular
376 proliferation and differentiation (24). Mutations in the gene are associated with neuronal and
377 endocrine diseases. TLCD1 affects membrane assembly and regulates membrane composition
378 and tolerance to fatty acids (25). FAM222B has been associated with red blood cell regulation
379 and protein binding using gene ontology (26). WNT7B is part of a pathway for embryonic
380 developmental processes and is linked to carcinogenesis and cancer development, in particular
381 gastric cancer (27–29). Additional genes near the four loci can be found in **Supplemental**
382 **Figures 2-5**. Overall, we found no suggestive functional basis to implicate these loci as under
383 selection in Native Hawaiians.
384
385 We postulate that even though the NH map displayed elevated correlation of recombination
386 rates genome-wide with many of the populations in 1KGP due to its recent admixture, it still
387 accurately reflects the recombination events pertaining to the Polynesian ancestries as
388 evidenced by its high correlation and overlap of hotspots with the PNS map. Assuming that
389 much of the signals of positive selection should predate the recent admixture times over the
390 last 10 generations or so, a population-specific map like the NH map captured much of the
391 recombination landscape from the Polynesian ancestry component and should improve the
392 robustness of haplotype-based scans of selection such as iHS. Indigenous populations such as
393 the Native Hawaiians have been understudied with respect to their present-day medical
394 conditions and healthcare considerations (30). Their evolutionary or adaptive histories for

395 surviving the most geographically isolated habitats on the planet for centuries could also
396 contribute to the genetic causes of differences in disease risk between the Native Hawaiians
397 and other continental populations. However, studies of indigenous populations have also been
398 fraught with "just-so" stories that only superficially tied evolutionary hypotheses to apparent
399 disparity in diseases (10,31,32). The availability of a population-specific recombination maps
400 could help weed out false positive findings, thus making the resulting conclusions more robust
401 and less harmful in future studies.
402
403 Finally, there are a number of limitations to our study, largely driven by the scarcity of genomic
404 knowledge and data for Native Hawaiians. We chose to use LDhat to construct the
405 recombination maps here as the available genomic data for Native Hawaiians are largely
406 restricted to the MEGA array data. When WGS data become available for Native Hawaiians at
407 scale, we would be able to evaluate more deeply the statistical genetic applications presented
408 here. For instance, WGS data would allow us to evaluate the impact of recombination maps on
409 imputation more thoroughly across the genome and different functional annotation.
410 Additionally, WGS data would also allow us to model the demographic history of the Native
411 Hawaiians, thus enabling the creation of more accurate recombination maps using methods,
412 such as pyrho (2). On top of the improved accuracy by incorporating demographic history,
413 pyrho can also infer recombination maps using hundreds of individuals, allowing us to reduce
414 noise by including many more individuals. Many of the evaluations we have conducted here
415 would be worth revisiting when this map based on sequenced data become available.
416 Furthermore, LDhat infers the relative, population-scaled, recombination rates, ρ, rather than

absolute recombination rates, r. We presumed the differences in recombination landscapes observed across populations reflected differences in LD pattern due to the Hawaiian's special demographic history. We were unable to assess differences in recombination between Native Hawaiians and other populations due to, for example, PRDM9 motif usage variations (1). Nevertheless, the relative rates estimated within Native Hawaiians are robust and these maps can inform genetic analyses for this population, as we have evaluated here. Finally, we also acknowledge that the underlying genetics and genomics should not supplant current standards through self-identity or genealogical records for defining community memberships; there is one Native Hawaiian population that cannot be discretized through genomics. The implications of these and future genetic findings on Pacific Islander health must be viewed through the lens of the social determinants of health with the goals to improve inclusion and equitable benefit sharing with the indigenous communities (33,34).

429 # Materials and Methods

430 ## Study cohort and data

431 The primary genetic dataset used here is a subcohort of the Multiethnic cohort (MEC), which is
432 a prospective epidemiological cohort of >215,000 individuals established as a collaboration
433 between the University of Hawaiʻi and University of Southern California (7). We focused on the
434 3,940 Native Hawaiian (MEC-NH) individuals genotyped on the Illumina MEGA array as part of
435 the PAGE consortium (35). We additionally used the MEGA array data from 5,325 African
436 Americans (MEC-AA) from the same cohort for specific comparisons in the present study.
437 Detailed descriptions of sample processing and QC can be found in previous publications (9,18).
438 In short, we restricted our analysis to biallelic SNPs with positions found in 1KGP and genotyped
439 in greater than 95% of the individuals, leaving 1,326,678 starting SNPs in MEC-NH population.
440 We selected two sets of 96 MEC-NH and one set of 96 MEC-AA individuals from these two
441 datasets to construct an LD-based recombination map (below). We then used the remaining
442 individuals for evaluations of the maps in downstream statistical and population genetic
443 applications.
444
445 In addition to the primary datasets from the PAGE consortium, we had access to 307 MEC-NH
446 individuals previously genotyped on the Illumina MEGAex array for a study of obesity (36) and
447 453 MEC-NH individuals previously genotyped on the Illumina 660W array for study of breast
448 cancer (37). For a subset of these NH individuals, we also had targeted exon sequencing data
449 across 746 exons in 37 genes as part of a breast cancer study (38); we used these data to

450 evaluate the impact of recombination maps on genotype imputation. In total, 607 sequenced
451 individuals overlap with our MEC-NH data: 453 from the NHBC study, 149 from the PAGE study,
452 and 5 from the obesity and PAGE studies.
453

454 Native Hawaiian-specific recombination map using LDhat and IBDrecomb

455 Recombination maps were inferred using LDhat with data ascertained on the MEGA array (39).
456 We modeled the analytic pipeline using LDhat after previously published descriptions on similar
457 efforts to infer the recombination landscape in humans and primates (2,11,40). LDhat relies on
458 lookup tables to enable tractable computation. Because the creation of lookup tables is
459 computationally taxing, we limited our analysis to the largest table based on 192 haplotypes
460 that was provided with the software. For this reason, we generated two recombination maps:
461 one using 96 individuals most enriched with the Polynesian ancestries that were found
462 predominantly in Native Hawaiians (PNS, proportion of Polynesian ancestries were previously
463 estimated (9)), and one using 96 individuals randomly selected from our Native Hawaiian
464 samples reflecting admixtures in our study sample (NH) (9,18). These subsets were chosen to
465 generate maps representative of a population with predominantly Polynesian ancestries and
466 the current NH population, respectively. We limited our analysis to the autosomes.
467

468 We updated the genome build of our data to hg38 using Liftover (41) and then phased all
469 available individuals with Eagle (v2.4.1, without including a phased reference) using the default
470 omnibus map based on HapMap populations (14,15,42) as supplied by Eagle. As comparisons,
471 we extracted 96 individuals from selected 1KGP populations to construct the recombination

472 map. In the few cases where fewer than 96 individuals were available from a population in
473 1000 Genomes such as ASW, CDX, GBR, LWK, MSL, and MXL, we extracted the number of
474 individuals based on the next available precomputed LDhat table. For CEU, we also downloaded
475 the OMNI array data (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502) to help
476 benchmark our pipeline. Next, each chromosome was split into windows of 4000 SNPs with 200
477 overlapping SNPs between windows. Finally, per previous studies (6,11,40), the *interval*
478 program from LDhat was run with the following settings: 30 million iterations, sampling set to
479 every 15,000 iterations, 7.5 million iterations used for burn-in, and a block penalty of 5. The
480 resulting estimates were integrated by removing the distal half of each overlapping region for
481 each window and combining all windows for each chromosome.
482
483 The LDhat *interval* program estimates population-scaled recombination rate for a set of
484 individuals using a composite likelihood method (39). The sex-averaged recombination rate, r,
485 can be recovered by the relationship $\rho = 4N_e r$, where N_e is the effective population size.
486 Following previous protocol (6,43), we regressed the population-scaled estimates from LDhat
487 for each population on the sex-averaged recombination rate from deCODE in order to estimate
488 $4N_e$ (**Supplemental Table 5**). Regression was performed by aggregating corresponding rates in 5
489 Mb windows across the autosomes. After rescaling by the inferred $4N_e$ for each population, we
490 then compute the recombination rate, r, per locus in centimorgan per megabase (cM/Mb).
491
492 To benchmark our pipeline, we applied the pipeline to OMNI array genotypes for 96 randomly
493 selected CEU individuals from 1KGP and then compared the inferred map from our pipeline to

494 the publicly available map (downloaded from
495 ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_r
496 ates), which was originally also inferred from the OMNI array data.
497
498 We also inferred the recombination map using IBDrecomb (11), which uses the end points of
499 IBD segments to infer recombination rates for a given set of individuals. IBDrecomb has been
500 shown to perform similarly to LDhat while outperforming admixture-based methods on
501 simulated and real data. To infer the recombination map using IBDrecomb, we first called IBD
502 segments using Refined IBD (44) (version 17Jan20.102) with default settings on genotypes
503 phased using the omnibus map. We then run the *merge-ibd-segments* tool with 1 allowed error
504 and maximum distance of 0.5 cM. Finally, we infer maps at 10 kb scale using these merged
505 segments with the remaining parameters at default settings. Because a large number of IBD
506 segments are needed to have sufficient number of informative historical recombination events,
507 we used all individuals in our Native Hawaiian cohort to detect IBD segments and infer
508 recombination map using IBDrecomb.
509
510 The inferred maps using both LDhat and IBDrecomb are released on
511 https://github.com/bldinh/recombination_maps in hg38.
512

513 Evaluating the impact of population-specific recombination map on
514 downstream statistical and population genetic applications

515 With each constructed Native Hawaiian-specific recombination map, we evaluated its impact
516 on four distinct statistical and population genetic applications where information of the
517 recombination landscapes are usually required for inference: imputation, local ancestry
518 inference, IBD segment calling and relatedness inference, and genome scans of positive
519 selection. In each case, we compared the performance and efficacy between using a Native
520 Hawaiian-specific map ("*NH map"*) or the default omnibus map released by Eagle (*"Eagle*
521 *map"*). We generally used the NH map for evaluation as most analyses incorporated the
522 entirety of the admixed NH study samples that were not used in map constructions. In all cases,
523 we excluded the 192 individuals that were used to construct both Native Hawaiian-specific
524 maps from evaluation.
525

526 ## Imputation

527 We phased and imputed our Native Hawaiian cohort (removing individuals that were used to
528 construct the recombination map after phasing) using either the NH map or the omnibus Eagle
529 map with HGDP+1KGP individuals (N = 3942) from gnomAD v3.1 as the imputation reference
530 panel (16). Phasing was completed using Eagle (14,15) (v2.4.1) with the same reference, and
531 imputation was performed in-house using Minimac4 (version 1.0.2) (45–47). To evaluate the
532 imputation accuracy, we compared the imputed genotype to exon sequencing data available
533 for a subset of 607 MEC-NH individuals. Imputation accuracy was measured by the squared
534 Pearson correlation of the sequenced genotype to the imputed dosage for variants that are
535 found in both datasets. In total, we compare 1,248,599 genotypes across 607 individuals and
536 2057 SNPs.

537

538 ## Local ancestry inference

539 We modeled Native Hawaiian individuals with four ancestry components corresponding to
540 those found most prevalently in Africa, East Asia, Europe, and Polynesia (9). A total of 708, 800,
541 and 671, individuals were selected from the HGDP+1KGP (from gnomAD v3.1) representing
542 African, East Asian, and European ancestries (**Supplemental Table 6**). In addition, we added 176
543 MEC-NH individuals previously estimated to have more than 90% Polynesian ancestry to create
544 a set of reference individuals for local ancestry inference (9,18) (after removing any individuals
545 used in recombination map constructions). In total, 3,665 MEC-NH individuals not used as
546 ancestry reference or in map construction were available to assess local ancestry inference (LAI)
547 concordance. We pre-phased every MEC-NH individual in the study sample together with the
548 HGDP+1KGP reference individual before separating the reference and test sets for LAI with
549 RFMix (v2.03-r0) (17). Because recombination maps are also used during phasing, we
550 performed our evaluation in two main comparisons: (1) comparison of RFMIX inference using
551 the same initial pre-phased data using the NH map and (2) comparison of RFMIX inference
552 including separately phasing the data on different maps. The former compares the direct effect
553 a recombination map has on LAI and the latter encompasses the difference a recombination
554 map would make on an analysis pipeline.

555

556 To compute concordance of LAI, at each site output by RFMix and for each individual we
557 calculated the joint probabilities of the inferred ancestry between the recombination maps. We
558 compared these inferences in a similar manner to a previous local ancestry study (48): if an

559 individual has a segment inferred as belonging to ancestry A and ancestry B using map 1,
560 whereas inference at the same location as ancestry C and ancestry A using map 2, we compare
561 the inferred haplotypes so as to maximize their agreement. In this example, we calculate the
562 inference as both maps agreeing on haplotype A and disagreeing for the other. Our joint
563 probability calculation uses the marginal probabilities for this maximal haplotype combination.
564

565 ## Integrated haplotype score

566 To evaluate the impact of recombination maps on genomic scans for signatures of positive
567 natural selection, we randomly selected 150 individuals from our Native Hawaiian cohort and
568 calculated iHS values using the selscan program (49). These scores were then normalized using
569 the *norm* program included with selscan. We compare the normalized iHS results from our NH
570 recombination map and the omnibus Eagle map. We used a normalized z-score of |4| as the
571 threshold for declaring a positively selected locus.
572

573 ## IBD segment calling and inference of relatedness

574 Using genotypes phased previously for LAI (above), we called IBD segments using Refined IBD
575 (version 17Jan20.102) (44). We then merged nearby IBD segments that could be spuriously
576 broken up using the *merge-ibd-segments* tool provided by the authors of Refined IBD with the
577 recommended parameters (1 allowed error, less than 0.6 cM distance). Initial analysis showed
578 an elevated number of spurious IBD segments within both maps. For example, 70% of all
579 detected IBD segments on chromosome 16 of length 15-20 cM map to the region located

580 approximately at 15,000,000 to 17,000,000 bp. This region appears to have spuriously elevated
581 recombination rate estimates; upon closer examination, the spurious pileup of IBD segments is
582 due to genomic regions with very low variant coverage. We thus instituted a heuristic to create
583 analysis masks (**below**) to be applied to the recombination maps before continuing IBD and
584 relatedness analyses. We excluded the gap-filling step to reduce potential overestimations of
585 shared IBD and excluded portions of IBD segments overlapping masked regions (additionally
586 excluding them from the *length of genome* calculations that are specific to each map). We then
587 used the IBD segments to infer genetic relatedness between any pair individuals by the
588 following calculation: $\varphi = \frac{sum\ IBD\ segments}{4\ x\ length\ of\ genome}$, where φ is the kinship coefficient. We used KING
589 (50), an allele-frequency based method, to independently call genetic relatedness and
590 restricted our comparisons due to the impact of different recombination maps to individuals
591 estimated to be 3rd-degree relatives and higher (kinship estimate > 0.0442). Our comparisons
592 were restricted to the intersection of pairs of individuals with estimated shared IBD segments
593 on both maps.
594

595 ## Analysis masks for recombination maps

596 Upon closer inspection of the distribution of IBD segments generated using the default omnibus
597 map and the NH map, we found an increase of segments with length ranging between 15-30
598 cM. We noticed that genomic regions with sparse SNP density and the resulting elevated
599 inferred rate of recombination would lead to accumulations of spurious IBD segments. These
600 spurious accumulations of lengthy IBD segments had been previously noted (51), and if

601 unadjusted, will lead to erroneous inference of familial relationships. To avoid inaccurate
602 results in downstream applications of the recombination maps, we created a filter to detect
603 regions with conditions that may bias analyses. In tiling windows of 500 kb length (shifting in 50
604 kb increments), we flagged the region within the window to be considered for removal from
605 downstream analysis if the number of markers available is less than 15 (compared to an
606 average of 242.50 markers per 500 kb across the autosome for our Native Hawaiian cohort on
607 the MEGA array). Other thresholds were considered, but we defined our threshold to attain a
608 balance between detection of regions that are outliers and minimizing the total length of
609 regions filtered out. Specifically, for our IBD analyses, we effectively set the genetic distance
610 across these regions to 0 cM in order to dampen the inflation of inferred segment lengths and
611 counts seen in our initial analysis. In total, our masks span 160.94 Mb (5.60%) of the autosome,
612 and the coordinates for the Native Hawaiians on the MEGA array in hg38 can be found in
613 **Supplemental Table 2**.

Acknowledgements

We would like to thank the Native Hawaiian participants in the Multiethnic Cohort that are involved in this study. The Multiethnic Cohort was funded through grants from the National Cancer Institute (U01CA164973, P01CA168530) and National Human Genome Research Institute (U01HG007397). We also would like to thank University of Hawai‘i Cancer Center’s Native Hawaiian Community Advisory Board for reviewing the study proposal and providing comments to earlier versions of this manuscript. This study is supported by grants from the National Institute of General Medical Sciences (R35GM142783 to C.W.K.C.) and the National Human Genome Research Institute (F31HG012159 to B.L.D.). Computation for this work is supported by the University of Southern California’s Center for High-Performance Computing (https://hpcc.usc.edu).

Citations

1. Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, et al. The landscape of recombination in African Americans. Nature. 2011 Jul 20;476(7359):170–5.

2. Spence JP, Song YS. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. Science Advances. 2019 Oct 1;5(10):eaaw9206.

3. Wegmann D, Kessner DE, Veeramah KR, Mathias RA, Nicolae DL, Yanek LR, et al. Recombination rates in admixed individuals identified by ancestry-based inference. Nature Genetics. 2011 Sep;43(9):847–53.

4. Hassan S, Surakka I, Taskinen MR, Salomaa V, Palotie A, Wessman M, et al. High-resolution population-specific recombination rates and their effect on phasing and genotype imputation. Eur J Hum Genet. 2021 Apr;29(4):615–24.

5. van Eeden G, Uren C, Pless E, Mastoras M, van der Spuy GD, Tromp G, et al. The recombination landscape of the Khoe-San likely represents the upper limits of recombination divergence in humans. Genome Biology. 2022 Aug 9;23(1):172.

6. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.

7. 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 Feb 15;151(4):346–57.

8. U.S. Department of Commerce [Internet]. 2023 [cited 2023 May 16]. U.S. Census Bureau Releases Key Stats in Honor of 2023 Asian American, Native Hawaiian, and Pacific Islander Heritage Month. Available from: https://www.commerce.gov/news/blog/2023/05/us-census-bureau-releases-key-stats-honor-2023-asian-american-native-hawaiian-and

9. Lin M, Caberto C, Wan P, Li Y, Lum-Jones A, Tiirikainen M, et al. Population-specific reference panels are crucial for genetic analyses: an example of the CREBRF locus in Native Hawaiians. Hum Mol Genet. 2020 Aug 3;29(13):2275–84.

10. Chiang CWK. The Opportunities and Challenges of Integrating Population Histories Into Genetic Studies for Diverse Populations: A Motivating Example From Native Hawaiians. Front Genet. 2021 Sep 27;12:643883.

11. Zhou Y, Browning BL, Browning SR. Population-Specific Recombination Maps from Segments of Identity by Descent. The American Journal of Human Genetics. 2020 Jul 2;107(1):137–48.

657 12. Spence JP, Song YS. Inference and analysis of population-specific fine-scale recombination
658 maps across 26 diverse human populations. Sci Adv. 2019 Oct 16;5(10):eaaw9206.

659 13. Bhérer C, Campbell CL, Auton A. Refined genetic maps reveal sexual dimorphism in human
660 meiotic recombination at multiple scales. Nat Commun. 2017 Apr 25;8(1):14994.

661 14. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-
662 based phasing using the Haplotype Reference Consortium panel. Nature Genetics. 2016
663 Nov;48(11):1443–8.

664 15. Loh PR, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort.
665 Nat Genet. 2016 Jul;48(7):811–6.

666 16. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational
667 constraint spectrum quantified from variation in 141,456 humans. Nature. 2020
668 May;581(7809):434–43.

669 17. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: A Discriminative Modeling Approach
670 for Rapid and Robust Local-Ancestry Inference. Am J Hum Genet. 2013 Aug 8;93(2):278–88.

671 18. Sun H, Lin M, Russell EM, Minster RL, Chan TF, Dinh BL, et al. The impact of global and local
672 Polynesian genetic ancestry on complex traits in Native Hawaiians. PLOS Genetics. 2021 Feb
673 11;17(2):e1009273.

674 19. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the
675 Human Genome. PLoS Biol. 2006 Mar;4(3):e72.

676 20. Chiang CWK, Ralph P, Novembre J. Conflation of Short Identity-by-Descent Segments Bias
677 Their Inferred Length Distribution. G3 (Bethesda). 2016 May 3;6(5):1287–96.

678 21. Browning BL, Browning SR. Detecting identity by descent and estimating genotype error
679 rates in sequence data. Am J Hum Genet. 2013 Nov 7;93(5):840–51.

680 22. Taparra K, Miller RC, Deville C. Navigating Native Hawaiian and Pacific Islander Cancer
681 Disparities From a Cultural and Historical Perspective. JCO Oncology Practice. 2021
682 Mar;17(3):130–4.

683 23. Harris DN, Song W, Shetty AC, Levano KS, Cáceres O, Padilla C, et al. Evolutionary genomic
684 dynamics of Peruvians before, during, and after the Inca Empire. Proceedings of the
685 National Academy of Sciences. 2018 Jul 10;115(28):E6526–35.

686 24. Ehring K, Manikowski D, Goretzko J, Froese J, Gude F, Jakobs P, et al. Conserved cholesterol-
687 related activities of Dispatched 1 drive Sonic hedgehog shedding from the cell membrane. J
688 Cell Sci. 2021 Aug 19;135(5):jcs258672.

689 25. Ruiz M, Bodhicharla R, Svensk E, Devkota R, Busayavalasa K, Palmgren H, et al. Membrane
690 fluidity is regulated by the C. elegans transmembrane protein FLD-1 and its human
691 homologs TLCD1/2. Elife. 2018 Dec 4;7:e40686.

692 26. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the
693 human binary protein interactome. Nature. 2020 Apr;580(7803):402–8.

694 27. Goessling W, North TE, Loewer S, Lord AM, Lee S, Stoick-Cooper CL, et al. Genetic
695 Interaction of PGE2 and Wnt Signaling Regulates Developmental Specification of Stem Cells
696 and Regeneration. Cell. 2009 Mar 20;136(6):1136–47.

697 28. Kirikoshi H, Sekihara H, Katoh M. Molecular cloning and characterization of human WNT7B.
698 International Journal of Oncology. 2001 Oct 1;19(4):779–83.

699 29. Gao Q, Yang L, Shen A, Li Y, Li Y, Hu S, et al. A WNT7B-m6A-TCF7L2 positive feedback loop
700 promotes gastric cancer progression and metastasis. Sig Transduct Target Ther. 2021 Feb
701 2;6(1):1–3.

702 30. Taparra K. Pacific Islanders Searching for Inclusion in Medicine. JAMA Health Forum. 2021
703 Feb 1;2(2):e210153.

704 31. Gosling AL, Buckley HR, Matisoo-Smith E, Merriman TR. Pacific Populations, Metabolic
705 Disease and 'Just-So Stories': A Critique of the 'Thrifty Genotype' Hypothesis in Oceania.
706 Annals of Human Genetics. 2015;79(6):470–80.

707 32. Fox K, Rallapalli KL, Komor AC. Rewriting Human History and Empowering Indigenous
708 Communities with Genome Editing Tools. Genes (Basel). 2020 Jan 12;11(1):88.

709 33. Pineda E, Benavente R, Gimmen MY, DeVille NV, Taparra K. Cancer Disparities among Pacific
710 Islanders: A Review of Sociocultural Determinants of Health in the Micronesian Region.
711 Cancers (Basel). 2023 Feb 22;15(5):1392.

712 34. Fox K. The Illusion of Inclusion — The "All of Us" Research Program and Indigenous Peoples'
713 DNA. N Engl J Med. 2020 Jul 30;383(5):411–3.

714 35. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of
715 diverse populations improves discovery for complex traits. Nature. 2019
716 Jun;570(7762):514–8.

717 36. Lim U, Monroe KR, Buchthal S, Fan B, Cheng I, Kristal BS, et al. Propensity for Intra-
718 abdominal and Hepatic Adiposity Varies Among Ethnic Groups. Gastroenterology. 2019
719 Mar;156(4):966-975.e10.

720 37. Siddiq A, Couch FJ, Chen GK, Lindström S, Eccles D, Millikan RC, et al. A meta-analysis of
721 genome-wide association studies of breast cancer identifies two novel susceptibility loci at
722 6q14 and 20q11. Hum Mol Genet. 2012 Dec 15;21(24):5373–84.

723 38. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J, et al. A Population-Based Study of
724 Genes Previously Implicated in Breast Cancer. N Engl J Med. 2021 Feb 4;384(5):440–51.

725 39. Auton A, McVean G. Recombination rate estimation in the presence of hotspots. Genome
726 Res. 2007 Aug;17(8):1219–27.

727 40. Xue C, Rustagi N, Liu X, Raveendran M, Harris RA, Venkata MG, et al. Reduced meiotic
728 recombination in rhesus macaques and the origin of the human recombination landscape.
729 PLOS ONE. 2020 Aug 25;15(8):e0236285.

730 41. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC
731 Genome Browser Database: update 2006. Nucleic Acids Res. 2006 Jan 1;34(Database
732 issue):D590-598.

733 42. International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et
734 al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007 Oct
735 18;449(7164):851–61.

736 43. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-
737 scale recombination rate differences between sexes, populations and individuals. Nature.
738 2010 Oct 28;467(7319):1099–103.

739 44. Browning BL, Browning SR. Improving the Accuracy and Efficiency of Identity-by-Descent
740 Detection in Population Data. Genetics. 2013 Jun 1;194(2):459–71.

741 45. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype
742 imputation service and methods. Nat Genet. 2016 Oct;48(10):1284–7.

743 46. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation.
744 Bioinformatics. 2015 Mar 1;31(5):782–4.

745 47. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype
746 imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012 Jul
747 22;44(8):955–9.

748 48. Browning SR, Grinde K, Plantinga A, Gogarten SM, Stilp AM, Kaplan RC, et al. Local Ancestry
749 Inference in a Large US-Based Hispanic/Latino Study: Hispanic Community Health
750 Study/Study of Latinos (HCHS/SOL). G3 (Bethesda). 2016 Mar 22;6(6):1525–34.

751 49. Szpiech ZA, Hernandez RD. selscan: An Efficient Multithreaded Program to Perform EHH-
752 Based Scans for Positive Selection. Mol Biol Evol. 2014 Oct;31(10):2824–7.

753 50. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship
754 inference in genome-wide association studies. Bioinformatics. 2010 Nov 15;26(22):2867–
755 73.

51. Ramstetter MD, Dyer TD, Lehman DM, Curran JE, Duggirala R, Blangero J, et al. Benchmarking Relatedness Inference Methods with Genome-Wide Data from Thousands of Relatives. Genetics. 2017 Sep;207(1):75–82.