Recombination map tailored to Native Hawaiians improves robustness of genomic scans for 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

Correspondence: charleston.chiang@med.usc.edu

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. Using suboptimal maps from distantly related populations may reduce the efficacy of genomic analyses, particularly for underrepresented populations such as the Native Hawaiians. To overcome this challenge, we constructed recombination maps using genome-wide array data from two study samples of Native Hawaiians: one reflecting the current admixed state of Native Hawaiians (NH map), and one based on individuals of enriched Polynesian ancestries (PNS map) with the 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 continental 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), likely driven by the unique demographic history of the Native Hawaiians. PNS also shared the fewest recombination hotspots with other populations (e.g. 8% of hotspots shared between PNS and CEU compared to 27% of hotspots shared between YRI and CEU). We found that downstream analyses 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, we found several loci with apparent genome-wide significant signals (|Z-score| > 4) in Native Hawaiians that would not have been significant when analyzed using NH-specific maps.

35 Population-specific recombination maps may therefore improve the robustness of haplotype-

36 based statistics and help us better characterize the evolutionary history that may underlie

37 Native 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 (Hassan et al., 2021; Hinch et al., 2011; Spence & Song, 2019a; van Eeden et al., 2022; Wegmann et al., 2011). 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 (Wegmann et al., 2011). The deCODE map, based on pedigrees of individuals with European-ancestries from Iceland, correlates better with CEU than YRI (Wegmann et al., 2011). 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 (Hinch et al., 2011). However, we have limited understanding of the recombination landscape for populations not represented by the 1000 Genomes Project (1KGP) (Auton et al., 2015), although efforts in developing these maps are starting for diverse populations such as the Nama people of southern Africa (van Eeden et al., 2022). Because large-scale pedigree data are rarely available for diverse populations, most attempts to characterize the recombination landscape of a population begin with population-scaled rates or their derivatives.

59 Given the prevalent use of recombination maps in genetic analysis and the fine-scale
60 differences between populations, it may be suboptimal to use a recombination map from a
61 distantly related population. This underdevelopment of available genomic resources, in general,
62 may further exacerbate the known disparity in transferring genomic insights and downstream
63 benefits to populations relatively distant from the study populations. Furthermore, the utility of
64 a population specific recombination map on downstream analyses has not been extensively
65 explored and studies have reached different qualitative conclusions. For an isolated,
66 homogenous population like the Finns (Hassan et al., 2021), a population-specific map did not
67 show a large impact on downstream analyses such as phasing and imputation, although this
68 could be due to general haplotypic similarity between Finns and the abundance of European-
69 ancestry samples represented in 1000 Genomes or that information from recombination plays
70 only a secondary role in the specific analyses evaluated. On the other hand, population-specific
71 recombination maps have also been shown to impact haplotype-based scans of positive
72 selection in the Nama (van Eeden et al., 2022), a population not adequately represented by
73 available maps which emphasize West African ancestry.
74
75 In this work, we focused on characterizing the recombination landscape for the Native
76 Hawaiians, using genome-wide genotyped individuals from the Multiethnic Cohort (MEC) Study
77 (Kolonel et al., 2000). Native Hawaiians are an admixed, indigenous, and underrepresented
78 population that, with a current population of 680,353 individuals, account for only 0.2% of the
79 U.S. population (*U.S. Census Bureau Releases Key Stats in Honor of 2023 Asian American, Native*
80 *Hawaiian, and Pacific Islander Heritage Month*, 2023). The lack of resources to advance

genomic research among Native Hawaiians has previously been detailed (Chiang, 2021; Lin et al., 2020). We thus also explored the impact of a population-specific map on multiple downstream analyses for the Native Hawaiians. The lack of available whole genome sequencing data and the relatively small cohort size limited the applicable methods that could be used to accurately infer a recombination for this population. We used LDhat (Auton & McVean, 2007) to estimate the recombination landscape. LDhat models the observed linkage disequilibrium (LD) between pairs of SNPs through a Bayesian reversible-jump Markov chain Monte Carlo process to estimate the population recombination rate across the genome and has been shown to infer accurate recombination maps using array data with good concordance to maps inferred through other methods (Spence & Song, 2019b; Zhou et al., 2020). Because of European colonization and subsequent waves of immigrations to the Hawaiian islands, we modeled the Native Hawaiians to include African-, East Asian-, and European-related ancestries in addition to Polynesian ancestries following previous work on this population (Lin et al., 2020; Sun et al., 2021). As a result, 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, and one based on a subset of individuals with enriched Polynesian ancestries. The latter map was constructed to gain insights on the recombination landscape of ancestral Polynesians and potentially provide a viable map for relatively unadmixed Polynesian populations such as the Samoans. We then evaluated the impact of a Native Hawaiian-specific map for downstream analyses such as LAI, imputation, IBD segment inference, and genome-wide scans of adaptation using haplotype-based statistics.

Materials and Methods

Study cohort and data

The primary genetic dataset is a subcohort of the Multiethnic Cohort (MEC) Study, which is a prospective epidemiological cohort of >215,000 individuals established as a collaboration between the University of Hawaiʻi and University of Southern California (Kolonel et al., 2000). We focused on the 3,940 Native Hawaiian (MEC-NH) individuals genotyped on the Illumina MEGA array as part of the PAGE consortium (Wojcik et al., 2019). We additionally used MEGA array data from 5,325 African Americans (MEC-AA) from the same cohort for specific comparisons in the present study. Detailed descriptions of sample processing and QC can be found in previous publications (Lin et al., 2020; Sun et al., 2021). In short, we restricted our analysis to biallelic SNPs with positions found in 1KGP and genotyped in greater than 95% of the individuals, leaving 1,326,678 starting SNPs in MEC-NH population. We selected two sets of 96 MEC-NH and one set of 96 MEC-AA individuals from these two datasets to construct an LD-based recombination map (below). We then used the remaining individuals for evaluations of the maps in downstream statistical and population genetic applications.

In addition to the primary datasets from the PAGE consortium, we had access to 307 MEC-NH individuals previously genotyped on the Illumina MEGAex array for a study of obesity (Lim et al., 2019) and 453 MEC-NH individuals previously genotyped on the Illumina 660W array for study of breast cancer (Siddiq et al., 2012). For a subset of these NH individuals, we also had targeted exon sequencing data across 746 exons in 37 genes as part of a breast cancer study

123 (Hu et al., 2021); we used these in combination to evaluate the impact of recombination maps
124 on genotype imputation. In total, 607 sequenced individuals overlap with our MEC-NH data:
125 453 from the NHBC study, 149 from the PAGE study, and 5 from the obesity and PAGE studies.
126

127 ## Native Hawaiian-specific recombination map using LDhat and IBDrecomb

128 Recombination maps were inferred using LDhat (Auton & McVean, 2007) with data ascertained
129 on the MEGA array. We modeled the analytic pipeline using LDhat after previously published
130 descriptions on similar efforts to infer the recombination landscape in humans and primates
131 (Spence & Song, 2019a; Xue et al., 2020; Zhou et al., 2020). LDhat relies on lookup tables to
132 allow for tractable computation. Because the creation of lookup tables is computationally
133 taxing, we limited our analysis to the largest table based on 192 haplotypes that was provided
134 with the software. For this reason, we generated two recombination maps: one using 96
135 individuals most enriched with the Polynesian ancestries that were found predominantly in
136 Native Hawaiians (PNS, proportion of Polynesian ancestries were previously estimated (Lin et
137 al., 2020)), and one using 96 individuals randomly selected from our Native Hawaiian samples
138 reflecting admixtures in our study sample (NH) (Lin et al., 2020; Sun et al., 2021). These subsets
139 were chosen to generate maps representative of a population with predominantly Polynesian
140 ancestries and the current NH population, respectively. We limited our analysis to the
141 autosomes.
142

143 We updated the genome build of our data to hg38 using Liftover (Hinrichs et al., 2006) and then
144 phased all available individuals with Eagle (v2.4.1, without including a phased reference) using

145 the default omnibus map based on HapMap populations (International HapMap Consortium et
146 al., 2007; Loh, Danecek, et al., 2016; Loh, Palamara, et al., 2016) as supplied by Eagle. As
147 comparisons, we extracted 96 individuals from selected 1KGP populations to construct the
148 recombination map. In the few cases where fewer than 96 individuals were available from a
149 population in 1000 Genomes such as ASW, CDX, GBR, LWK, MSL, and MXL, we extracted the
150 number of individuals based on the next available precomputed LDhat table. For CEU, we also
151 downloaded OMNI array data (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502) to
152 help benchmark our pipeline. Next, each chromosome was split into windows of 4000 SNPs
153 with 200 overlapping SNPs between windows. Finally, per previous studies (Auton et al., 2015;
154 Xue et al., 2020; Zhou et al., 2020), the *interval* program from LDhat was run with the following
155 settings: 30 million iterations, sampling set to every 15,000 iterations, 7.5 million iterations
156 used for burn-in, and a block penalty of 5. The resulting estimates were integrated by removing
157 the distal half of each overlapping region for each window and combining all windows for each
158 chromosome.
159
160 The LDhat *interval* program estimates population-scaled recombination rate for a set of
161 individuals using a composite likelihood method (Auton & McVean, 2007). The sex-averaged
162 recombination rate, r, can be recovered by the relationship $\rho = 4N_e r$, where N_e is the effective
163 population size. Following previous protocol (Auton et al., 2015; Kong et al., 2010), we
164 regressed the population-scaled estimates from LDhat for each population on the sex-averaged
165 recombination rate from deCODE in order to estimate $4N_e$ (**Supplemental Table 1**). Regression
166 was performed by aggregating corresponding rates in 5 Mb windows across the autosomes.

167 After rescaling by the inferred $4N_e$ for each population, we then compute the recombination
168 rate, r, per locus in centimorgan per megabase (cM/Mb).
169
170 To benchmark our pipeline, we applied it to OMNI array genotypes for 96 randomly selected
171 CEU individuals from 1KGP and then compared the inferred map from our pipeline to the
172 publicly available map (downloaded from
173 ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_r
174 ates), which was originally also inferred from the OMNI array data.
175
176 We also inferred a recombination map for NH using IBDrecomb (Zhou et al., 2020), which uses
177 the end points of IBD segments to infer recombination rates for a given set of individuals.
178 IBDrecomb has been shown to perform similarly to LDhat while outperforming admixture-
179 based methods on simulated and real data. To infer the recombination map using IBDrecomb,
180 we first called IBD segments using Refined IBD (B. L. Browning & Browning, 2013b) (version
181 17Jan20.102) with default settings on genotypes phased using the omnibus map. We then run
182 the *merge-ibd-segments* tool with 1 allowed error and maximum distance of 0.5 cM. Finally, we
183 infer maps at 10 kb scale using these merged segments with the remaining parameters at
184 default settings. Because a large number of IBD segments are needed to have sufficient number
185 of informative historical recombination events, we used all individuals in our Native Hawaiian
186 cohort to detect IBD segments before inferring a recombination map with IBDrecomb.
187

188 The inferred maps using both LDhat and IBDrecomb are released on
189 https://github.com/bldinh/recombination_maps in hg38.
190

191 Evaluating the impact of population-specific recombination map on
192 downstream statistical and population genetic applications

193 With each constructed Native Hawaiian-specific recombination map, we evaluated its impact
194 on four distinct statistical and population genetic applications where information of the
195 recombination landscapes are usually required for inference: imputation, local ancestry
196 inference, IBD segment calling and relatedness inference, and genome scans of positive
197 selection. In each case, we compared the performance and efficacy between using a Native
198 Hawaiian-specific map (NH map) or the default omnibus map released by Eagle (Eagle map).
199 We generally used the NH map for evaluation as most analyses incorporated the entirety of the
200 admixed NH study samples that were not used in map constructions. In all cases, we excluded
201 the 192 individuals that were used to construct both Native Hawaiian-specific maps from
202 evaluation.
203

204 Imputation

205 We phased and imputed our Native Hawaiian cohort (removing individuals that were used to
206 construct the recombination map after phasing) using either the NH map or the omnibus Eagle
207 map with HGDP+1KGP individuals (N = 3942) from gnomAD v3.1 as the imputation reference
208 panel (Karczewski et al., 2020). Phasing was completed using Eagle v2.4.1 (Loh, Danecek, et al.,

209 2016; Loh, Palamara, et al., 2016) with the same reference, and imputation was performed in-
210 house using Minimac4 version 1.0.2 (Das et al., 2016; Fuchsberger et al., 2015; Howie et al.,
211 2012). To evaluate the imputation accuracy, we compared the imputed genotype to exon
212 sequencing data available for a subset of 607 MEC-NH individuals. Imputation accuracy was
213 measured by the squared Pearson correlation of the sequenced genotype to the imputed
214 dosage for variants that are found in both datasets. In total, we compare 1,248,599 genotypes
215 across 607 individuals and 2057 SNPs.
216

217 ## Local ancestry inference

218 We modeled Native Hawaiian individuals with four ancestry components corresponding to
219 those found most prevalently in Africa, East Asia, Europe, and Polynesia (Lin et al., 2020). A
220 total of 708, 800, and 671, individuals were selected from the HGDP+1KGP (from gnomAD v3.1)
221 representing African, East Asian, and European ancestries (**Supplemental Table 2**). In addition,
222 we added 176 MEC-NH individuals previously estimated to have more than 90% Polynesian
223 ancestry to create a set of reference individuals for local ancestry inference (Lin et al., 2020;
224 Sun et al., 2021) (after removing any individuals used in recombination map constructions). In
225 total, 3,665 MEC-NH individuals not used as ancestry reference or in map construction were
226 available to assess local ancestry inference (LAI) concordance. We pre-phased every MEC-NH
227 individual in the study sample together with the HGDP+1KGP reference individual before
228 separating the reference and test sets for LAI with RFMIX v2.03-r0 (Maples et al., 2013).
229 Because recombination maps are also used during phasing, we performed our evaluation in two
230 main comparisons: (1) comparison of RFMIX inference using the same initial pre-phased data

using the NH map and (2) comparison of RFMIX inference after separately phasing the data on different maps. The former compares the direct effect a recombination map has on LAI and the latter encompasses the difference a recombination map would make on an analysis pipeline.

To compute concordance of LAI, at each site output by RFMix and for each individual we calculated the joint probabilities of the inferred ancestry between the recombination maps. We compared these inferences in a similar manner to a previous local ancestry study (S. R. Browning et al., 2016): if an individual has a segment inferred as belonging to ancestry A and ancestry B using map 1 and from ancestry C and ancestry A using map 2, we compare the inferred haplotypes so as to maximize their agreement. In this example, we calculate the inference as both maps agreeing on haplotype A and disagreeing for the other. Our joint probability calculation uses the marginal probabilities for this maximal haplotype combination.

Integrated haplotype score

To evaluate the impact of recombination maps on genomic scans for signatures of positive natural selection, we randomly selected 150 individuals from our Native Hawaiian cohort and calculated iHS values using the selscan program (Szpiech & Hernandez, 2014). These scores were then normalized using the *norm* program included with selscan. We compared the normalized iHS results from our NH recombination map and two alternatives that would be commonly used in practice: the omnibus Eagle map and the pedigree-based European map from deCODE. We used a normalized z-score of |4| as the threshold for declaring a putative

252 positively selected locus using either the omnibus or the pedigree map and examined the
253 corresponding score on the NH map.
254

255 IBD segment calling and inference of relatedness

256 Using genotypes phased previously for LAI (above), we called IBD segments using Refined IBD
257 version 17Jan20.102 (B. L. Browning & Browning, 2013b). We then merged nearby IBD
258 segments that could be spuriously broken up using the *merge-ibd-segments* tool provided by
259 the authors of Refined IBD with the recommended parameters (1 allowed error, less than 0.6
260 cM distance). Initial analysis showed an elevated number of spurious IBD segments within both
261 maps. For example, 70% of all detected IBD segments on chromosome 16 of length 15-20 cM
262 map to the region located approximately at 15,000,000 to 17,000,000 bp. This region appears
263 to have spuriously elevated recombination rate estimates; upon closer examination, the
264 spurious pileup of IBD segments is due to genomic regions with very low variant coverage. We
265 thus instituted a heuristic to create analysis masks (**below**) to be applied to the recombination
266 maps before continuing IBD and relatedness analyses. We excluded the gap-filling step to
267 reduce potential overestimations of shared IBD and excluded portions of IBD segments
268 overlapping masked regions (additionally excluding them from the *length of genome*
269 calculations that are specific to each map). We then used the IBD segments to infer genetic
270 relatedness between any pair individuals by the following calculation: $\varphi = \frac{sum\ IBD\ segments}{4\ x\ length\ of\ genome}$,
271 where φ is the kinship coefficient. We used KING (Manichaikul et al., 2010), an allele-frequency
272 based method, to independently call genetic relatedness and restricted our comparisons due to

273 the impact of different recombination maps to individuals estimated to be 3rd-degree relatives
274 and higher (kinship estimate > 0.0442). Our comparisons were restricted to the intersection of
275 pairs of individuals with estimated shared IBD segments on both maps.
276

277 Analysis masks for recombination maps

278 Upon closer inspection of the distribution of IBD segments generated using the default omnibus
279 map and the NH map, we found an increase of segments with length ranging between 15-30
280 cM. We noticed that genomic regions with sparse SNP density and the resulting elevated
281 inferred rate of recombination would lead to accumulations of spurious IBD segments. These
282 spurious accumulations of lengthy IBD segments had been previously noted (Ramstetter et al.,
283 2017), and if unadjusted, will lead to erroneous inference of familial relationships. To avoid
284 inaccurate results in downstream applications of the recombination maps, we created a filter to
285 detect regions with conditions that may bias analyses. In tiling windows of 500 kb length
286 (shifting in 50 kb increments), we flagged the region within the window to be considered for
287 removal from downstream analysis if the number of markers available is less than 15
288 (compared to an average of 242.50 markers per 500 kb across the autosome for our Native
289 Hawaiian cohort on the MEGA array). Other thresholds were considered, but we defined our
290 threshold to attain a balance between detection of regions that are outliers and minimizing the
291 total length of regions filtered out. Specifically, for our IBD analyses, we effectively set the
292 genetic distance across these regions to 0 cM in order to dampen the inflation of inferred
293 segment lengths and counts seen in our initial analysis. In total, our masks span 160.94 Mb

294 (5.60%) of the autosome, and the coordinates for the Native Hawaiians on the MEGA array in

295 hg38 can be found in **Supplemental Table 3**.

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. Population-level rates, and the LD information used in inferring these rates, can be influenced by past population sizes. Therefore, following previous practices, we inferred and regressed out the contribution of the demographic history as measured by effective population size, N_e (**Materials and Methods**), in order to better compare the inferred recombination landscape across populations (but also see **Discussion** for limitations). 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 was 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 (Bhérer et al., 2017). 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 concluded that our implemented pipeline near-faithfully recreated the previously published map, and that, given the same SNP content, differences in the distribution

316 of any genotyping errors attributed to data generation platform (OMNI array vs. low pass

317 sequencing) did not substantially impact the recombination map inference.

318

319 **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

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

331

332 Because the Native Hawaiian genetic data from the MEC is only available on the MEGA array,

333 we also benchmarked the impact using a different set of SNPs on map inference. Again, using

334 the CEU 1KGP population as a model, we extracted the genotypes found on the MEGA array to

335 infer a recombination map. The resulting map is highly correlated with the version generated

336 based on extracted OMNI array SNPs, at least for 50 kb scale or higher ($r > 0.91$; **Table 1**). Taken

337 together, our LDhat inference pipeline, when applied to SNPs found on the MEGA array for the

338 CEU population, would produce highly concordant recombination maps at 50 kb scale or higher

339 compared to the published map. Therefore, the recombination landscape for Native Hawaiians

340 that we inferred here can be compared to those published for other 1KGP populations.
341 Nevertheless, for between-population comparisons of the recombination landscape we
342 compared the inferred NH maps to those based on the MEGA array SNP content extracted from
343 1KGP sequence data.
344

345 Contrasting the Native Hawaiian recombination landscape

346 We proceeded to infer a recombination map for 96 randomly selected admixed Native
347 Hawaiians from our study sample (NH map) to represent the current Native Hawaiian
348 population and a map for 96 individuals from our study sample previously estimated to be most
349 enriched with the Polynesian ancestries found predominantly in Native Hawaiians (PNS map;
350 **Materials and Methods**). The PNS map was constructed to provide insights on the
351 recombination landscape of the ancient Polynesians and how it may differ with other
352 continental populations around the world. However, we caution that the Polynesian ancestries
353 found predominantly in Native Hawaiians may or may not have diverged from those found in
354 other Polynesian populations such as the Samoans. Consistent with their known history of serial
355 bottleneck and long-term isolation (Chiang, 2021), the inferred past population size,
356 parameterized by the effective population size (N_e), was substantially smaller for PNS than
357 other populations in 1KGP (**Supplemental Table 1**). The estimated N_e for NH was larger, more in
358 line with other continental and admixed populations in 1KGP. Regressing out the effect of
359 demographic history (**Materials and Methods**), we then compared the inferred Native
360 Hawaiian recombination landscape with that from other populations that were similarly
361 constructed using MEGA array genotypes extracted from the 1KGP phase 3 datasets.

362

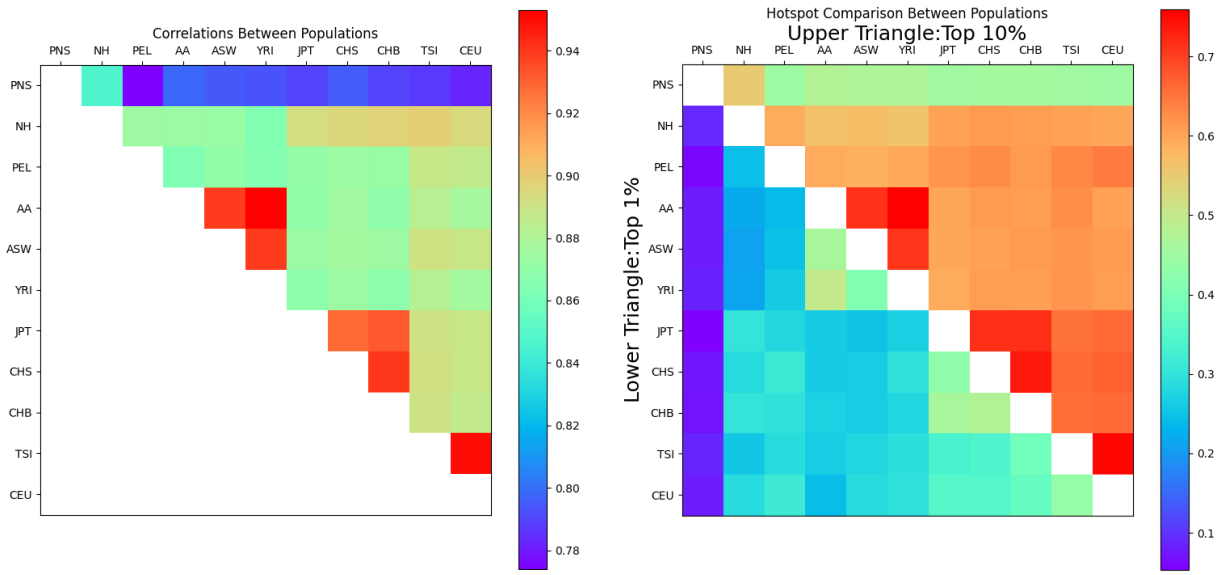
363 We found that populations with recent shared history displayed higher levels of correlation, as
364 expected (**Figure 1a**). This is notable for populations representing each of the ancestral
365 components found: African (MEC-AA, ASW, and YRI), East Asian (CHB, CHS, and JPT), and
366 European (CEU, TSI). At the 50 kb scale, the recombination landscape as indicated by the NH
367 map showed intermediate correlations with the non-Native Hawaiian populations (r = 0.865 -
368 0.899), tending to correlate best with populations representing East Asian (r = 0.892 – 0.898)
369 and European ancestry (r = 0.895 - 0.899), likely due to substantial admixture from these two
370 ancestral components. In contrast, the landscape from PNS showed the lowest correlations
371 with almost all populations compared (r = 0.774 - 0.847). PNS correlated most strongly with NH
372 (r = 0.847) and most weakly with Peruvian in Lima, Peru (PEL; r = 0.774). Overall, PNS correlated
373 more poorly with other populations than PEL, a population with a known history of isolation.
374 Taken together, our results are consistent with PNS representing a previously unrepresented
375 component of ancestry by 1KGP and may suggest a stronger effect of isolation or enhanced
376 genetic drift in the past compared to PEL (although we also note that we did not prune PEL to
377 enrich for Indigenous American ancestry due to small sample size).

378

379 These trends were also observed when we compared recombination hotspot sharing between
380 populations (**Figure 1b**). Following methodology of previous studies (Wegmann et al., 2011), we
381 divided the genome into 50 kb windows and defined hotspots for each population as the
382 windows in the top 1 percentile or 10 percentile with the highest average recombination rate.
383 For each pair of populations, we then compared the shared proportion of hotspots. The

relationships at both hotspot thresholds were similar to the ones observed for correlations of the genome-wide recombination landscape (**Figure 1a**): populations expected to represent recent shared history had the highest sharing at both thresholds. In contrast, PNS had lower hotspot sharing relative to other populations except for NH, with the lowest sharing with PEL. Compared to PNS, the NH map has better relative sharing with the other 1KGP populations, including slightly higher sharing with European and East Asian populations. PNS's low correlation genome-wide in recombination rate variations and hotspot sharing with other populations is thus indicative of a unique recombination landscape.

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



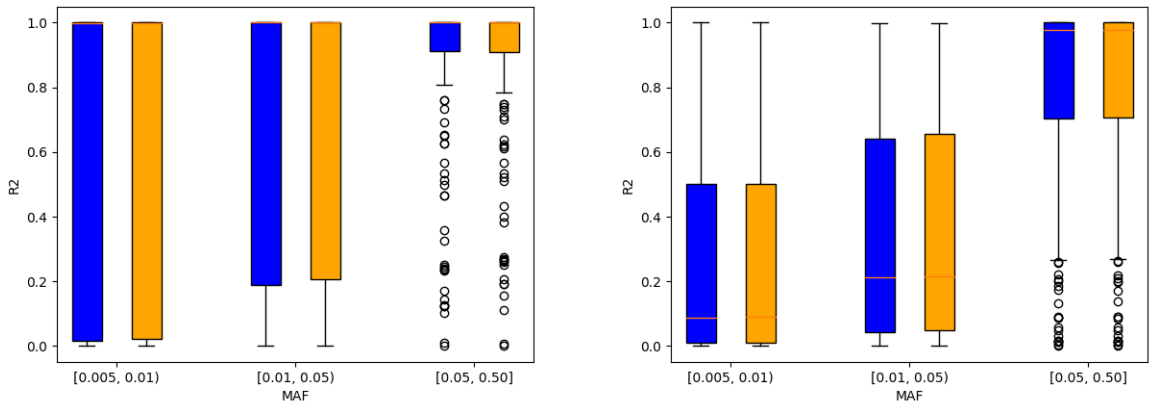
Correlations between populations at 50 kb scale (**A**) and sharing of recombination hotspots defined as the top 1% (lower left) or 10% (upper right) windows 50 kb in size by inferred recombination rate (**B**). Both the correlations and proportion of sharing were calculated by excluding genomic regions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

403 ## Impact of the recombination map on imputation

404 We evaluated the impact of a population-specific map for Native Hawaiians on downstream
405 statistical and population genetic applications, specifically focusing on the impact of the
406 recombination map on genotype imputation, local ancestry inference, IBD and relatedness
407 inference, and genomic scans for positive selection.
408
409 Imputation is a common analysis that uses haplotype sharing to infer genotypes at variants not
410 directly observed in a given dataset. The recombination map is needed for phasing (a pre-
411 requisite for imputation) and for imputation itself as it provides the prior that the ancestral
412 haplotype may have switched from one haplotype to another at a given genomic location. We
413 thus evaluated the impact of using a potentially mis-specified omnibus map (*i.e.* the default
414 recombination map released by the commonly used phasing software, Eagle (Loh, Danecek, et
415 al., 2016; Loh, Palamara, et al., 2016)) in phasing and imputation, compared to the population-
416 specific map constructed in the present study.
417
418 We imputed 453 and 154 Native Hawaiian individuals genotyped on the Illumina Human660W
419 (NHBC) and MEGA arrays, respectively, against the 1KGP+HGDP reference panel released by
420 gnomAD (Karczewski et al., 2020). Compared to the targeted exon sequencing data we have on
421 overlapping individuals, we were able to evaluate imputation accuracy in the 658 and 482 SNPs
422 for NHBC and MEGA cohorts, respectively. Across three minor allele frequency (MAF) bins, 0.5 –
423 1%, 1 – 5%, and 5 – 50%, we observed no notable difference in imputation accuracy when

imputation was performed using the NH map compared to the omnibus map (**Figure 2**, **Supplemental Table 4**)

Figure 2: evaluation of the impact of recombination maps on imputation accuracy



Comparison of imputation accuracy using the NH map (blue) and the omnibus map (orange). MAF for each SNP was computed using the sequencing data. **Left**: 154 individuals on the MEGA array were compared on 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 (Maples et al., 2013), 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

445 strongly. These results are also consistent with previous findings that even a naïve uniform

446 recombination map would produce <10% of discordance in inferred ancestry calls (Sun et al.,

447 2021).

448

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

450 **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%

451

452 **Phasing and LAI with different maps**

Omnibus map \ NH map	PNS	AFR	EAS	EUR
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%

453 The Native Hawaiian population is modeled as a four-way admixture, with ancestry components from
454 Polynesia (PNS), Africa (AFR), East Asia (EAS), and Europe (EUR).
455

456 Impact of recombination map on IBD segment calling and relatedness

457 inference

458 Our initial IBD analysis identified regions of the genome where we observed extreme pileups of

459 IBD segments that are spurious (>90% of individuals would share IBD segments locally). We

460 thus created analysis masks to remove these spurious regions driven by insufficient SNP

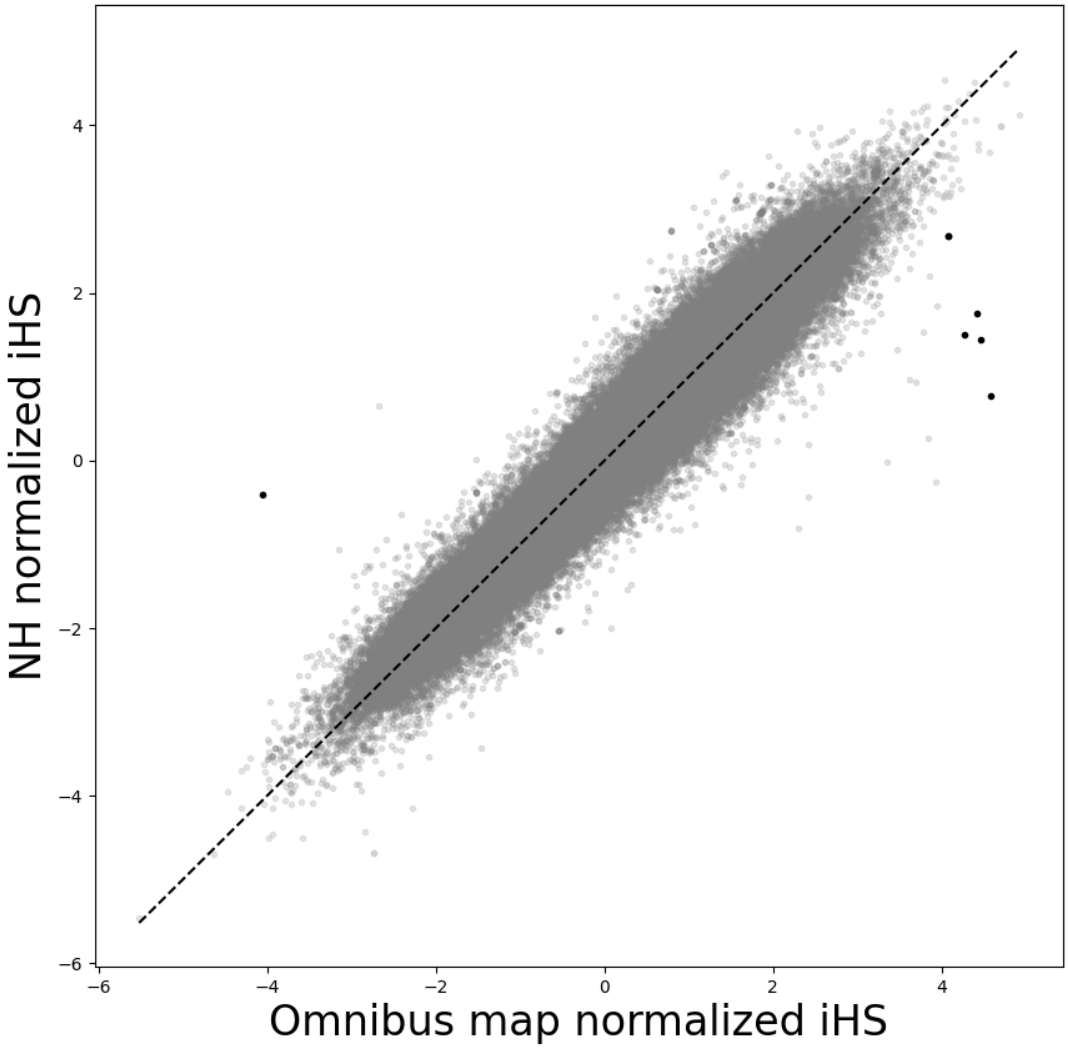
461 coverage (see **Materials and Methods**, **Supplemental Figure 1**). When comparing the

462 distribution of IBD segments using masked versions of both maps, we observed little difference
463 (**Supplemental Figure 1**). Given that the inferred recombination rates in regions of sparse SNP
464 density could be unreliable and confound downstream analysis such as IBD segment calling, we
465 also released a set of masked regions (**Supplemental Table 3**) for users to apply when using
466 these recombination maps.
467
468 We used the kinship coefficient (φ) based on the pairwise proportion of IBD sharing to compare
469 differences in relatedness inference due to the recombination map (**Materials and Methods**).
470 Of the 218,132 pairs of individuals inferred by KING to be 3rd-degree or closer relatives, 200,617
471 were found in both the omnibus and NH maps to compare. The estimated kinship coefficient
472 from each map was calculated independently and the maps were found to be highly
473 concordant with a Pearson correlation coefficient of 0.995.
474

475 Impact of recombination map on genomic scans for positive selection
476 using integrated haplotype score

477 Lastly, we evaluated the impact of recombination maps on the robustness of a haplotype-based
478 metric of positive selection, the integrated haplotype score (iHS) (Voight et al., 2006). We
479 considered loci with normalized iHS absolute Z-score (|Z|) > 4 as candidate loci under positive
480 selection. Most SNPs in our scans using different recombination maps were concordant with
481 similar iHS values (**Figure 3**). However, we identified 6 SNPs across 4 loci (nearby genes: DISP1,
482 TLCD1, FAM222B, and WNT7B) that showed |Z| values > 4 when using the omnibus map but

483 normalized |Z| values between 0 to 2 when using the NH map (**Supplemental Figures 2-5**).
484 Moreover, there are additional loci that would have attenuated evidence of selection in
485 analysis using the NH map (|Z|< 2) but fall short of our conservative cutoff of 4 standard
486 deviations when using the omnibus map. Selection scans may also be performed using a
487 pedigree-based map, such as one from deCODE, to avoid any confounding between haplotype-
488 based statistics and local LD patterns. We also compared the analysis using the deCODE
489 pedigree map (**Supplemental Figure 6**), which is based on a European-ancestry population in
490 Iceland. We observed 46 outlier SNPs across 19 loci that would have attenuated evidence of
491 selection (|Z| > 4 using deCODE map, but < 2 using NH map). Due to the substantial increase of
492 candidate loci under positive selection with the deCODE map and the overall larger iHS after
493 normalization (the largest |Z| near 15), it is clear that haplotype-based selection analyses are
494 sensitive to the choice of recombination maps. Therefore, using the population-specific
495 recombination map in haplotype-based selection scans may better protect against spurious
496 detections of loci under positive selection.
497
498 **Figure 3: Genomic scan for positive selection using haplotype-based statistics is sensitive to**
499 **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

511 have sufficient number of IBD segments for inference, we inferred a recombination map using
512 3,937 individuals using IBDrecomb (Zhou et al., 2020), which would be more akin to the NH map
513 constructed using LDhat. The resulting IBD-based map showed large concordance with the NH
514 map (**Supplemental Table 5**), although the correlation may be lower compared to that reported
515 previously for African Americans (Zhou et al., 2020). For instance, the correlation between an
516 IBD-based map of an African American cohort and an LDhat map from ASW was previously
517 reported to be as high as 0.95 at the 500kb scale; we observed a correlation of 0.90, 0.90, and
518 0.92 at the 500 kb, 1 Mb, and 5 Mb scales (**Supplemental Table 5**). This may be due to the
519 differences in the precision in IBD segment detection using array versus sequencing data (B. L.
520 Browning & Browning, 2013a; Chiang et al., 2016), which may percolate to the precision in the
521 constructed recombination maps, or it may be due to differences in ancestry composition
522 between our African American cohort (sampled largely in Southern California) with that tested
523 previously (Zhou et al., 2020). In comparison, the correlation between IBD-based map and
524 LDhat map from NH are between 0.85 to 0.89 across the same scales (**Supplemental Table 5**).
525 Nevertheless, we found that compared to the omnibus map, the IBD-based map also did not
526 substantially impact the downstream inference for imputation (**Supplemental Figure 7**), LAI
527 (**Supplemental Table 6**; 94.9% concordance, comparing IBD-based to LDhat-based map), and
528 kinship coefficient estimates (Pearson correlation with NH map: 0.9945). We also found that
529 genomic scan for positive selection using iHS would be more robust if we were to use an IBD-
530 based recombination map for NH as well (**Supplemental Figure 8**), confirming our previous
531 observation that a population-specific recombination map would be important for performing
532 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 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

554 the Native Hawaiians also contributed to the health disparity experienced by the population
555 today, and the genetic risks to diseases must be evaluated considering this history (Sun et al.,
556 2021; Taparra et al., 2021). As such, we also constructed the latter map, to better reflect the
557 genomic composition of the Native Hawaiians today. The NH map would be more suited for
558 genetic epidemiologic studies that use a population-based sample and is the primary map that
559 we evaluated throughout this study.
560
561 Overall, we found that the recombination landscape from the PNS map stood out to be the
562 most unique compared to other 1KGP and MEC populations based on the correlation of
563 recombination rate variations across the genome and proportion of hotspot sharing. Notably,
564 the pairwise sharing between YRI and CEU, two populations known to have divergent
565 population histories, have higher correlation with one another (r = 0.876) than either
566 population with PNS (0.793 and 0.783 for YRI and CEU, respectively). The landscape based on
567 the NH map, in contrast, showed higher correlations with the maps from other populations,
568 with slightly higher correlation with the maps inferred from European and East Asian
569 populations. This is likely due to the recent admixture from other continental populations seen
570 within the NH cohort today (Sun et al., 2021). Lastly, the PNS map showed the lowest
571 correlation with PEL, and both PNS and PEL maps showed relatively low correlations with maps
572 from other populations examined here. The Peruvian from Lima, Peru, is known for its history
573 of isolation relative to other 1KGP populations (Harris et al., 2018). The observation here thus
574 suggests that long-term isolation and lower effective population sizes are driving the LD pattern
575 and shaping the recombination landscape. The low correlation between PNS and PEL, and

576 between the two with other populations, are potentially reflecting their respective
577 independent isolation histories.
578
579 We evaluated the impact of a population-specific map on downstream population and
580 statistical genetic analyses. In general, we found little difference due to the choice of the
581 recombination maps, particularly for imputation accuracy, local ancestry inference, and IBD
582 segment detections and genetic relatedness inference. In local ancestry inference, we did
583 observe an increased discordance between the two maps, if the input data were also phased
584 separately using different maps (**Table 2**), suggesting that the choice of recombination maps
585 would have a small but tangible impact on these downstream analyses. However, in each of
586 these applications, the recombination maps are used mostly as priors while the actual inference
587 is ultimately driven by the genetic data. Therefore, even though there are fine-scale differences
588 in the recombination landscape between Native Hawaiians and other 1KGP populations, the
589 accuracy of these downstream applications will not be substantially impacted by the choice of
590 the map used.
591
592 Our findings suggest that the choice of the recombination map could make qualitative
593 differences in population genetic analyses such as genomic scans for positive selection using
594 the integrated haplotype score (iHS) because these statistics would incorporate recombination
595 information in their calculation, representative of haplotype length surrounding a SNP. We
596 found that 6 SNPs across 4 loci appeared to be putative selection loci (using a threshold of
597 $|iHS| > 4$ after normalization) but have attenuated signals using the NH map ($|iHS| < 2$). The

598 signals in the four loci are driven by the genes DISP1, TLCD1, FAM222B, and WNT7B. We
599 examined each locus manually and did not find clear evidence for them to be under selection.
600 DISP1 impacts signaling and transport for cellular proliferation and differentiation (Ehring et al.,
601 2021). Mutations in the gene are associated with neuronal and endocrine diseases. TLCD1
602 affects membrane assembly and regulates membrane composition and tolerance to fatty acids
603 (Ruiz et al., 2018). FAM222B has been associated with red blood cell regulation and protein
604 binding using gene ontology (Luck et al., 2020). WNT7B is part of a pathway for embryonic
605 developmental processes and is linked to carcinogenesis and cancer development, in particular
606 gastric cancer (Gao et al., 2021; Goessling et al., 2009; Kirikoshi et al., 2001). Additional genes
607 near the four loci can be found in **Supplemental Figures 2-5**. Overall, we found no suggestive
608 functional basis to implicate these loci as under selection in Native Hawaiians.
609
610 We postulate that even though the NH map displayed elevated correlation of recombination
611 rates genome-wide with many of the populations in 1KGP due to its recent admixture, it still
612 accurately reflects the recombination events pertaining to the Polynesian ancestries as
613 evidenced by its high correlation and overlap of hotspots with the PNS map. Assuming that
614 much of the signals of positive selection should predate the recent admixture times over the
615 last 10 generations or so, a population-specific map like the NH map captured much of the
616 recombination landscape from the Polynesian ancestry component and should improve the
617 robustness of haplotype-based scans of selection such as iHS. Indigenous populations such as
618 the Native Hawaiians have been understudied with respect to their present-day medical
619 conditions and healthcare considerations (Taparra, 2021). Their evolutionary or adaptive

histories for surviving the most geographically isolated habitats on the planet for centuries could also contribute to the genetic causes of differences in disease risk between the Native Hawaiians and other continental populations. However, studies of indigenous populations have also been fraught with “just-so” stories that only superficially tied evolutionary hypotheses to apparent disparity in diseases (Chiang, 2021; Fox et al., 2020; Gosling et al., 2015). The availability of a population-specific recombination maps could help weed out false positive findings, thus making the resulting conclusions more robust and less harmful in future studies.

Finally, there are a number of limitations to our study, largely driven by the scarcity of genomic knowledge and data for Native Hawaiians. We chose to use LDhat to construct the recombination maps here as the available genomic data for Native Hawaiians are largely restricted to the MEGA array data. When WGS data become available for Native Hawaiians at scale, we would be able to evaluate more deeply the statistical genetic applications presented here. For instance, WGS data would allow us to evaluate the impact of recombination maps on imputation more thoroughly across the genome and different functional annotation. Additionally, WGS data would also allow us to model the demographic history of the Native Hawaiians, thus enabling the creation of more accurate recombination maps using methods, such as pyrho (Spence & Song, 2019a). On top of the improved accuracy by incorporating demographic history, pyrho can also infer recombination maps using hundreds of individuals, allowing us to reduce noise by including many more individuals. Many of the evaluations we have conducted here would be worth revisiting when this map based on sequenced data become available. Furthermore, LDhat infers the relative, population-scaled, recombination

642 rates, ρ, rather than absolute recombination rates, r. Despite the effort to control for the
643 impact of population history through estimating and regressing out the long-term effective
644 population size, we presumed the differences in recombination landscapes observed across
645 populations largely reflected differences in LD pattern due to the Hawaiian's unique
646 demographic history. We were unable to assess differences in recombination between Native
647 Hawaiians and other populations due to, for example, PRDM9 motif usage variations (Hinch et
648 al., 2011). To directly and reliably estimate the underlying recombination rates would require
649 large-scale pedigree data, such as the one amassed by deCODE (Kong et al., 2010), but is
650 extremely difficult to ascertain for a vulnerable, indigenous population such as Native
651 Hawaiians. Nevertheless, the rates estimated by LDhat within Native Hawaiians are highly
652 correlated with IBD-based maps, and these maps can inform and improve genetic analyses
653 using haplotype-based statistics for this population, as we have evaluated here. This is thus an
654 important step towards inclusion and lessening further irresponsible construct of selection
655 stories for the Native Hawaiians. Finally, we also acknowledge that the underlying genetics and
656 genomics should not supplant current standards through self-identity or genealogical records
657 for defining community memberships; there is one Native Hawaiian population that cannot be
658 discretized through genomics. The implications of these and future genetic findings on Pacific
659 Islander health must be viewed through the lens of the social determinants of health with the
660 goals to improve inclusion and equitable benefit sharing with the indigenous communities (Fox,
661 2020; Pineda et al., 2023).

Acknowledgements

We would like to thank John Novembre and Vagheesh Narasimhan for discussions and comments on the preprinted version of this manuscript. We would also 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 Hawaiʻi 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

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., … National Eye Institute, N. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), Article 7571. https://doi.org/10.1038/nature15393

Auton, A., & McVean, G. (2007). Recombination rate estimation in the presence of hotspots. *Genome Research*, *17*(8), 1219–1227. https://doi.org/10.1101/gr.6386707

Bhérer, C., Campbell, C. L., & Auton, A. (2017). Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms14994

Browning, B. L., & Browning, S. R. (2013a). Detecting identity by descent and estimating genotype error rates in sequence data. *American Journal of Human Genetics*, *93*(5), 840–851. https://doi.org/10.1016/j.ajhg.2013.09.014

Browning, B. L., & Browning, S. R. (2013b). Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics*, *194*(2), 459–471. https://doi.org/10.1534/genetics.113.150029

Browning, S. R., Grinde, K., Plantinga, A., Gogarten, S. M., Stilp, A. M., Kaplan, R. C., Avilés-Santa, M. L., Browning, B. L., & Laurie, C. C. (2016). Local Ancestry Inference in a Large US-Based Hispanic/Latino Study: Hispanic Community Health Study/Study of Latinos

694 (HCHS/SOL). *G3: Genes|Genomes|Genetics*, *6*(6), 1525–1534.
695 https://doi.org/10.1534/g3.116.028779
696 Chiang, C. W. K. (2021). The Opportunities and Challenges of Integrating Population Histories
697 Into Genetic Studies for Diverse Populations: A Motivating Example From Native
698 Hawaiians. *Frontiers in Genetics*, *12*, 643883.
699 https://doi.org/10.3389/fgene.2021.643883
700 Chiang, C. W. K., Ralph, P., & Novembre, J. (2016). Conflation of Short Identity-by-Descent
701 Segments Bias Their Inferred Length Distribution. *G3 (Bethesda, Md.)*, *6*(5), 1287–1296.
702 https://doi.org/10.1534/g3.116.027581
703 Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy,
704 S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W. G., Swaroop, A.,
705 Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-
706 generation genotype imputation service and methods. *Nature Genetics*, *48*(10), Article
707 10. https://doi.org/10.1038/ng.3656
708 Ehring, K., Manikowski, D., Goretzko, J., Froese, J., Gude, F., Jakobs, P., Rescher, U., Kirchhefer,
709 U., & Grobe, K. (2021). Conserved cholesterol-related activities of Dispatched 1 drive
710 Sonic hedgehog shedding from the cell membrane. *Journal of Cell Science*, *135*(5),
711 jcs258672. https://doi.org/10.1242/jcs.258672
712 Fox, K. (2020). The Illusion of Inclusion—The "All of Us" Research Program and Indigenous
713 Peoples' DNA. *New England Journal of Medicine*, *383*(5), 411–413.
714 https://doi.org/10.1056/NEJMp1915987

715 Fox, K., Rallapalli, K. L., & Komor, A. C. (2020). Rewriting Human History and Empowering
716 Indigenous Communities with Genome Editing Tools. *Genes*, *11*(1), 88.
717 https://doi.org/10.3390/genes11010088
718 Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2015). minimac2: Faster genotype imputation.
719 *Bioinformatics*, *31*(5), 782–784. https://doi.org/10.1093/bioinformatics/btu704
720 Gao, Q., Yang, L., Shen, A., Li, Y., Li, Y., Hu, S., Yang, R., Wang, X., Yao, X., & Shen, G. (2021). A
721 WNT7B-m6A-TCF7L2 positive feedback loop promotes gastric cancer progression and
722 metastasis. *Signal Transduction and Targeted Therapy*, *6*(1), Article 1.
723 https://doi.org/10.1038/s41392-020-00397-z
724 Goessling, W., North, T. E., Loewer, S., Lord, A. M., Lee, S., Stoick-Cooper, C. L., Weidinger, G.,
725 Puder, M., Daley, G. Q., Moon, R. T., & Zon, L. I. (2009). Genetic Interaction of PGE2 and
726 Wnt Signaling Regulates Developmental Specification of Stem Cells and Regeneration.
727 *Cell*, *136*(6), 1136–1147. https://doi.org/10.1016/j.cell.2009.01.015
728 Gosling, A. L., Buckley, H. R., Matisoo-Smith, E., & Merriman, T. R. (2015). Pacific Populations,
729 Metabolic Disease and 'Just-So Stories': A Critique of the 'Thrifty Genotype' Hypothesis
730 in Oceania. *Annals of Human Genetics*, *79*(6), 470–480.
731 https://doi.org/10.1111/ahg.12132
732 Harris, D. N., Song, W., Shetty, A. C., Levano, K. S., Cáceres, O., Padilla, C., Borda, V., Tarazona,
733 D., Trujillo, O., Sanchez, C., Kessler, M. D., Galarza, M., Capristano, S., Montejo, H.,
734 Flores-Villanueva, P. O., Tarazona-Santos, E., O'Connor, T. D., & Guio, H. (2018).
735 Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire.

Proceedings of the National Academy of Sciences, *115*(28), E6526–E6535. https://doi.org/10.1073/pnas.1720798115

Hassan, S., Surakka, I., Taskinen, M.-R., Salomaa, V., Palotie, A., Wessman, M., Tukiainen, T., Pirinen, M., Palta, P., & Ripatti, S. (2021). High-resolution population-specific recombination rates and their effect on phasing and genotype imputation. *European Journal of Human Genetics*, *29*(4), Article 4. https://doi.org/10.1038/s41431-020-00768-8

Hinch, A. G., Tandon, A., Patterson, N., Song, Y., Rohland, N., Palmer, C. D., Chen, G. K., Wang, K., Buxbaum, S. G., Akylbekova, M., Aldrich, M. C., Ambrosone, C. B., Amos, C., Bandera, E. V., Berndt, S. I., Bernstein, L., Blot, W. J., Bock, C. H., Boerwinkle, E., … Myers, S. R. (2011). The landscape of recombination in African Americans. *Nature*, *476*(7359), 170–175. https://doi.org/10.1038/nature10336

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T. S., Harte, R. A., Hsu, F., Hillman-Jackson, J., Kuhn, R. M., Pedersen, J. S., Pohl, A., Raney, B. J., Rosenbloom, K. R., Siepel, A., Smith, K. E., Sugnet, C. W., … Kent, W. J. (2006). The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Research*, *34*(Database issue), D590-598. https://doi.org/10.1093/nar/gkj144

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*(8), 955–959. https://doi.org/10.1038/ng.2354

Hu, C., Hart, S. N., Gnanaolivu, R., Huang, H., Lee, K. Y., Na, J., Gao, C., Lilyquist, J., Yadav, S., Boddicker, N. J., Samara, R., Klebba, J., Ambrosone, C. B., Anton-Culver, H., Auer, P.,

758 Bandera, E. V., Bernstein, L., Bertrand, K. A., Burnside, E. S., … Couch, F. J. (2021). A
759 Population-Based Study of Genes Previously Implicated in Breast Cancer. *The New*
760 *England Journal of Medicine*, *384*(5), 440–451.
761 https://doi.org/10.1056/NEJMoa2005936
762 International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve,
763 L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S.,
764 Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., … Stewart, J.
765 (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*,
766 *449*(7164), 851–861. https://doi.org/10.1038/nature06258
767 Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L.,
768 Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M.,
769 Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., …
770 MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation
771 in 141,456 humans. *Nature*, *581*(7809), Article 7809. https://doi.org/10.1038/s41586-
772 020-2308-7
773 Kirikoshi, H., Sekihara, H., & Katoh, M. (2001). Molecular cloning and characterization of human
774 WNT7B. *International Journal of Oncology*, *19*(4), 779–783.
775 https://doi.org/10.3892/ijo.19.4.779
776 Kolonel, L. N., Henderson, B. E., Hankin, J. H., Nomura, A. M., Wilkens, L. R., Pike, M. C., Stram,
777 D. O., Monroe, K. R., Earle, M. E., & Nagamine, F. S. (2000). A multiethnic cohort in
778 Hawaii and Los Angeles: Baseline characteristics. *American Journal of Epidemiology*,
779 *151*(4), 346–357. https://doi.org/10.1093/oxfordjournals.aje.a010213

780 Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., Sigurdsson, A., Jonasdottir, A.,
781 Walters, G. B., Jonasdottir, A., Gylfason, A., Kristinsson, K. T., Gudjonsson, S. A., Frigge,
782 M. L., Helgason, A., Thorsteinsdottir, U., & Stefansson, K. (2010). Fine-scale
783 recombination rate differences between sexes, populations and individuals. *Nature*,
784 *467*(7319), 1099–1103. https://doi.org/10.1038/nature09525
785 Lim, U., Monroe, K. R., Buchthal, S., Fan, B., Cheng, I., Kristal, B. S., Lampe, J. W., Hullar, M. A.,
786 Franke, A. A., Stram, D. O., Wilkens, L. R., Shepherd, J., Ernst, T., & Marchand, L. L.
787 (2019). Propensity for Intra-abdominal and Hepatic Adiposity Varies Among Ethnic
788 Groups. *Gastroenterology*, *156*(4), 966-975.e10.
789 https://doi.org/10.1053/j.gastro.2018.11.021
790 Lin, M., Caberto, C., Wan, P., Li, Y., Lum-Jones, A., Tiirikainen, M., Pooler, L., Nakamura, B.,
791 Sheng, X., Porcel, J., Lim, U., Setiawan, V. W., Le Marchand, L., Wilkens, L. R., Haiman, C.
792 A., Cheng, I., & Chiang, C. W. K. (2020). Population-specific reference panels are crucial
793 for genetic analyses: An example of the CREBRF locus in Native Hawaiians. *Human*
794 *Molecular Genetics*, *29*(13), 2275–2284. https://doi.org/10.1093/hmg/ddaa083
795 Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K.,
796 Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., & Price, A. L. (2016).
797 Reference-based phasing using the Haplotype Reference Consortium panel. *Nature*
798 *Genetics*, *48*(11), Article 11. https://doi.org/10.1038/ng.3679
799 Loh, P.-R., Palamara, P. F., & Price, A. L. (2016). Fast and accurate long-range phasing in a UK
800 Biobank cohort. *Nature Genetics*, *48*(7), Article 7. https://doi.org/10.1038/ng.3571

801 Luck, K., Kim, D.-K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T.,
802 Campos-Laborie, F. J., Charloteaux, B., Choi, D., Coté, A. G., Daley, M., Deimling, S.,
803 Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., … Calderwood, M. A.
804 (2020). A reference map of the human binary protein interactome. *Nature*, *580*(7803),
805 Article 7803. https://doi.org/10.1038/s41586-020-2188-x
806 Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust
807 relationship inference in genome-wide association studies. *Bioinformatics*, *26*(22),
808 2867–2873. https://doi.org/10.1093/bioinformatics/btq559
809 Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: A Discriminative
810 Modeling Approach for Rapid and Robust Local-Ancestry Inference. *American Journal of*
811 *Human Genetics*, *93*(2), 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020
812 Pineda, E., Benavente, R., Gimmen, M. Y., DeVille, N. V., & Taparra, K. (2023). Cancer Disparities
813 among Pacific Islanders: A Review of Sociocultural Determinants of Health in the
814 Micronesian Region. *Cancers*, *15*(5), 1392. https://doi.org/10.3390/cancers15051392
815 Ramstetter, M. D., Dyer, T. D., Lehman, D. M., Curran, J. E., Duggirala, R., Blangero, J., Mezey, J.
816 G., & Williams, A. L. (2017). Benchmarking Relatedness Inference Methods with
817 Genome-Wide Data from Thousands of Relatives. *Genetics*, *207*(1), 75–82.
818 https://doi.org/10.1534/genetics.117.1122
819 Ruiz, M., Bodhicharla, R., Svensk, E., Devkota, R., Busayavalasa, K., Palmgren, H., Ståhlman, M.,
820 Boren, J., & Pilon, M. (2018). Membrane fluidity is regulated by the C. elegans
821 transmembrane protein FLD-1 and its human homologs TLCD1/2. *ELife*, *7*, e40686.
822 https://doi.org/10.7554/eLife.40686

Siddiq, A., Couch, F. J., Chen, G. K., Lindström, S., Eccles, D., Millikan, R. C., Michailidou, K., Stram, D. O., Beckmann, L., Rhie, S. K., Ambrosone, C. B., Aittomäki, K., Amiano, P., Apicella, C., Baglietto, L., Bandera, E. V., Beckmann, M. W., Berg, C. D., Bernstein, L., … Vachon, C. M. (2012). A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Human Molecular Genetics*, *21*(24), 5373–5384. https://doi.org/10.1093/hmg/dds381

Spence, J. P., & Song, Y. S. (2019a). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances*, *5*(10), eaaw9206. https://doi.org/10.1126/sciadv.aaw9206

Spence, J. P., & Song, Y. S. (2019b). Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Science Advances*, *5*(10), eaaw9206. https://doi.org/10.1126/sciadv.aaw9206

Sun, H., Lin, M., Russell, E. M., Minster, R. L., Chan, T. F., Dinh, B. L., Naseri, T., Reupena, M. S., Lum-Jones, A., the Samoan Obesity, L., Cheng, I., Wilkens, L. R., Marchand, L. L., Haiman, C. A., & Chiang, C. W. K. (2021). The impact of global and local Polynesian genetic ancestry on complex traits in Native Hawaiians. *PLOS Genetics*, *17*(2), e1009273. https://doi.org/10.1371/journal.pgen.1009273

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827. https://doi.org/10.1093/molbev/msu211

Taparra, K. (2021). Pacific Islanders Searching for Inclusion in Medicine. *JAMA Health Forum*, *2*(2), e210153. https://doi.org/10.1001/jamahealthforum.2021.0153

845 Taparra, K., Miller, R. C., & Deville, C. (2021). Navigating Native Hawaiian and Pacific Islander
846 Cancer Disparities From a Cultural and Historical Perspective. *JCO Oncology Practice*,
847 *17*(3), 130–134. https://doi.org/10.1200/OP.20.00831
848 *U.S. Census Bureau Releases Key Stats in Honor of 2023 Asian American, Native Hawaiian, and*
849 *Pacific Islander Heritage Month*. (2023, May 1). U.S. Department of Commerce.
850 https://www.commerce.gov/news/blog/2023/05/us-census-bureau-releases-key-stats-
851 honor-2023-asian-american-native-hawaiian-and
852 van Eeden, G., Uren, C., Pless, E., Mastoras, M., van der Spuy, G. D., Tromp, G., Henn, B. M., &
853 Möller, M. (2022). The recombination landscape of the Khoe-San likely represents the
854 upper limits of recombination divergence in humans. *Genome Biology*, *23*(1), 172.
855 https://doi.org/10.1186/s13059-022-02744-5
856 Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A Map of Recent Positive
857 Selection in the Human Genome. *PLoS Biology*, *4*(3), e72.
858 https://doi.org/10.1371/journal.pbio.0040072
859 Wegmann, D., Kessner, D. E., Veeramah, K. R., Mathias, R. A., Nicolae, D. L., Yanek, L. R., Sun, Y.
860 V., Torgerson, D. G., Rafaels, N., Mosley, T., Becker, L. C., Ruczinski, I., Beaty, T. H.,
861 Kardia, S. L. R., Meyers, D. A., Barnes, K. C., Becker, D. M., Freimer, N. B., & Novembre, J.
862 (2011). Recombination rates in admixed individuals identified by ancestry-based
863 inference. *Nature Genetics*, *43*(9), Article 9. https://doi.org/10.1038/ng.894
864 Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., Highland, H. M.,
865 Patel, Y. M., Sorokin, E. P., Avery, C. L., Belbin, G. M., Bien, S. A., Cheng, I., Cullina, S.,
866 Hodonsky, C. J., Hu, Y., Huckins, L. M., Jeff, J., Justice, A. E., … Carlson, C. S. (2019).

Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, *570*(7762), Article 7762. https://doi.org/10.1038/s41586-019-1310-4

Xue, C., Rustagi, N., Liu, X., Raveendran, M., Harris, R. A., Venkata, M. G., Rogers, J., & Yu, F. (2020). Reduced meiotic recombination in rhesus macaques and the origin of the human recombination landscape. *PLOS ONE*, *15*(8), e0236285. https://doi.org/10.1371/journal.pone.0236285

Zhou, Y., Browning, B. L., & Browning, S. R. (2020). Population-Specific Recombination Maps from Segments of Identity by Descent. *The American Journal of Human Genetics*, *107*(1), 137–148. https://doi.org/10.1016/j.ajhg.2020.05.016