The American Journal of Human Genetics, Volume *109*

Supplemental information

The genomic signatures of natural selection in admixed human populations

Sebastian Cuadros-Espinoza, Guillaume Laval, Lluis Quintana-Murci, and Etienne Patin

Supplementary Figures

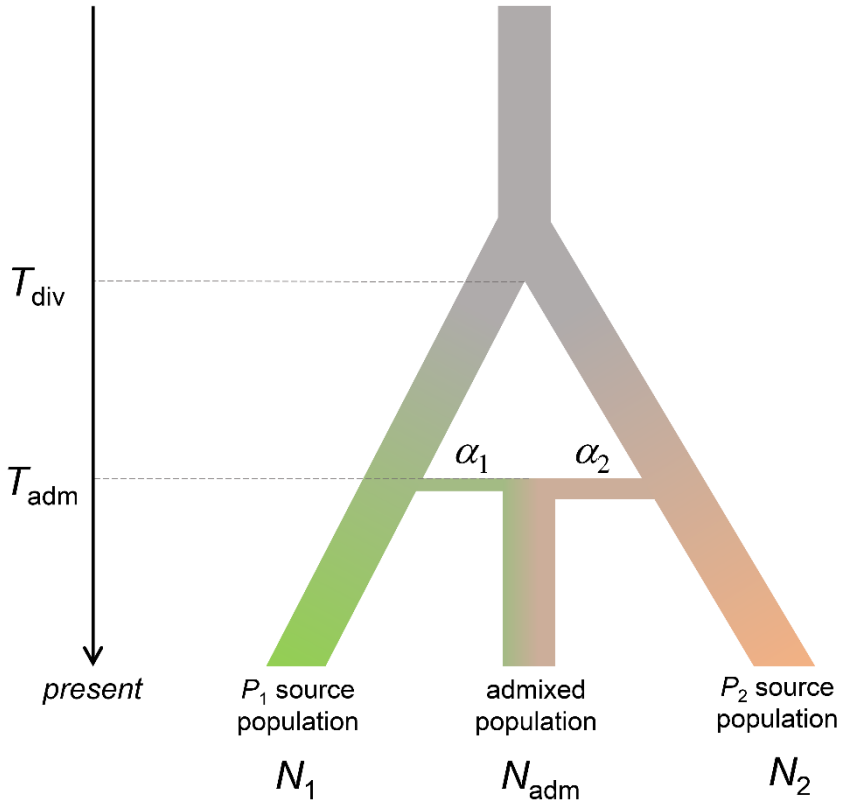

Figure S1. The simulated single-pulse admixture model.

The admixed population originates from admixture between two source populations, referred to as P_1 and P_2. P_1 and P_2 contribute α_1 and α_2 admixture proportions to the admixed population, with $\alpha_1 + \alpha_2 = 1$. P_1 and P_2 diverge T_{div} generations ago and the admixture event occurs T_{adm} generations ago. The population sizes of the admixed population and of P_1 and P_2 source populations are N_{adm}, N_1 and N_2, respectively.

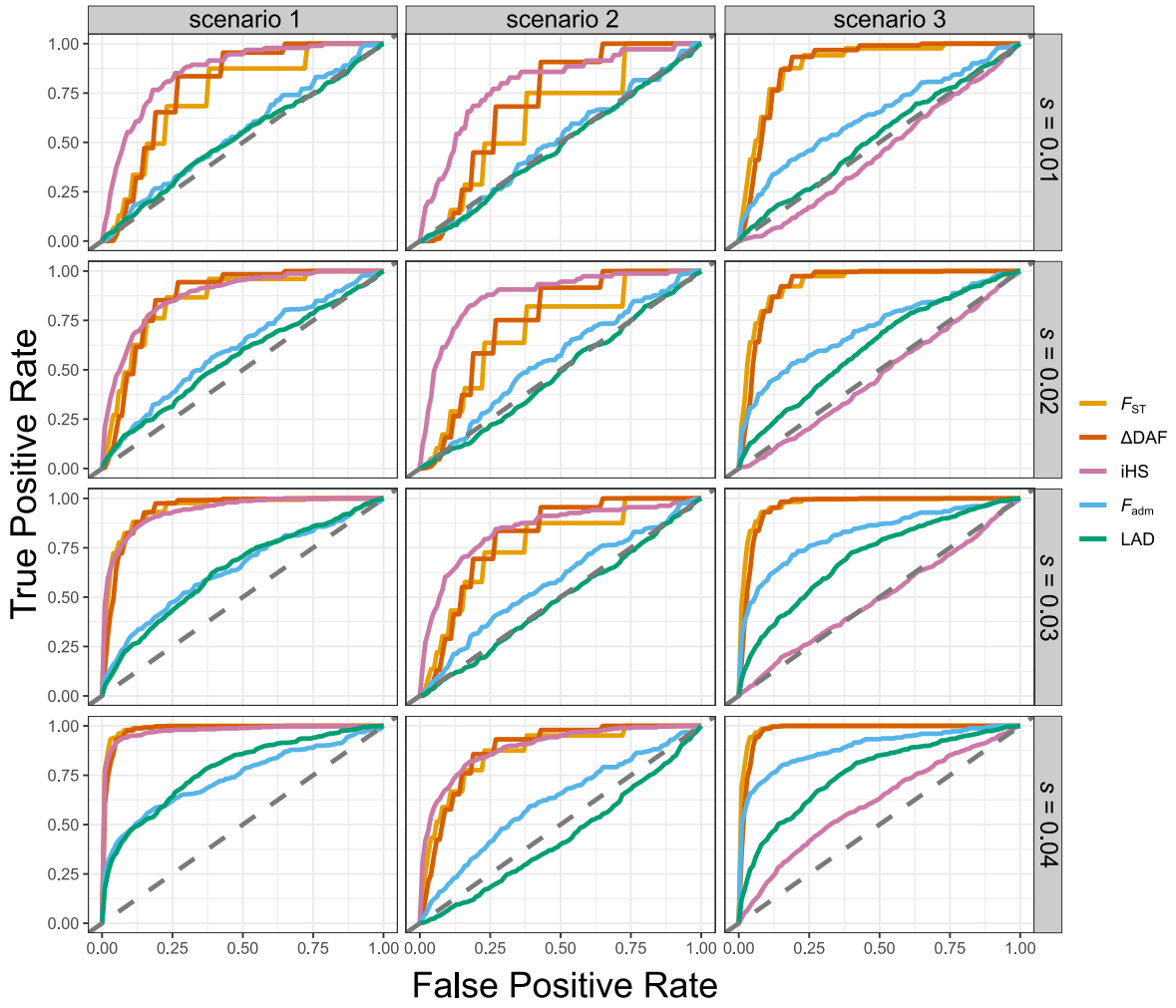

Figure S2. Performance of neutrality statistics under different scenarios of admixture with selection, assuming different selection coefficients.

Receiver operating characteristic (ROC) curves comparing the performance of the classic neutrality statistics F_{ST}, ΔDAF and iHS and the admixture-specific statistics F_{adm} and LAD, across the 3 explored admixture with selection scenarios, with varying selection coefficients $s \in \{0.01, 0.02, 0.03, 0.04\}$.

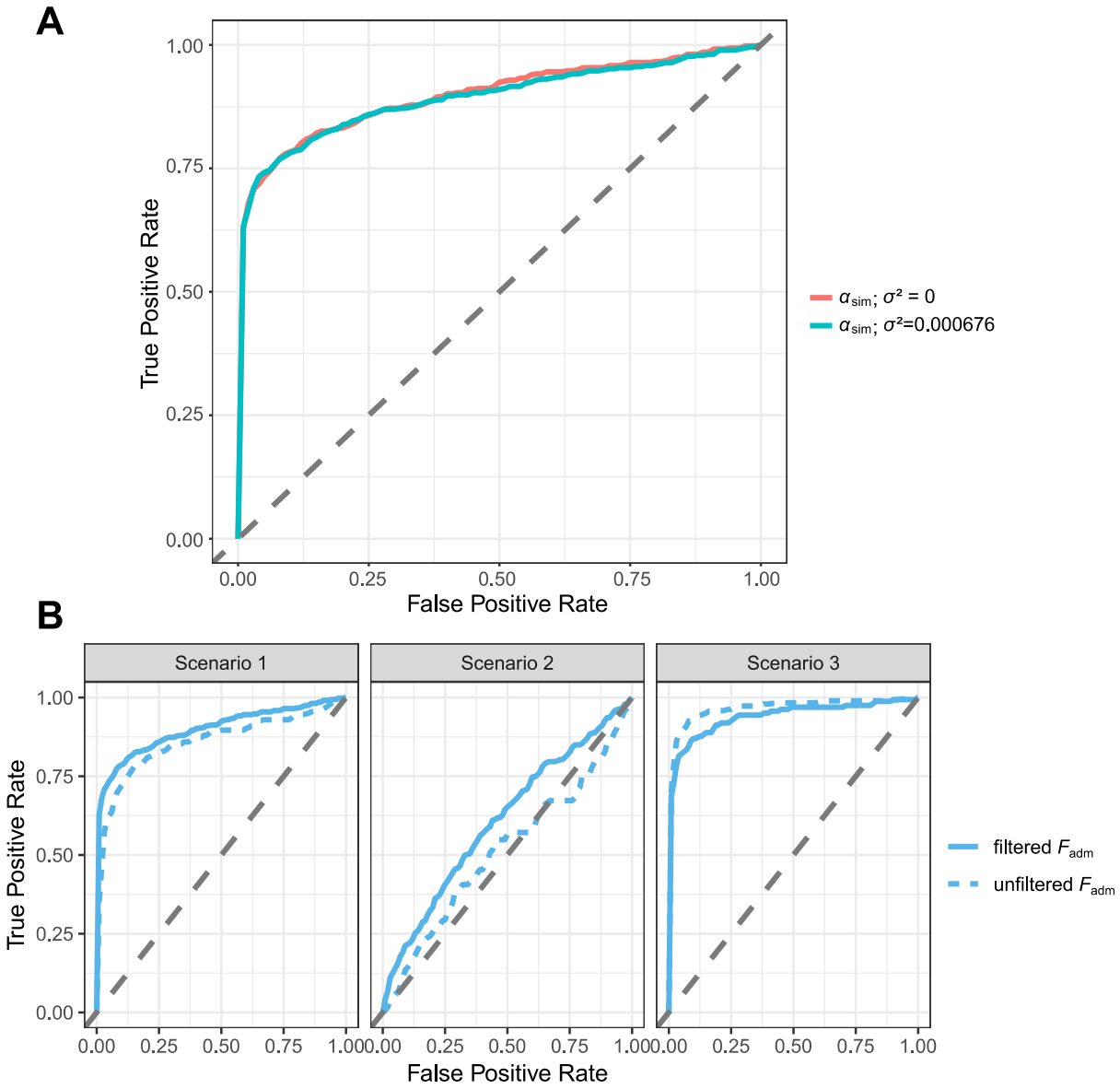

Figure S3. Performance of F_{adm} when using simulated admixture proportions with error and when applying or not an allele frequency filter.

(A) Receiver operating characteristic (ROC) curves comparing the performance of F_{adm} when using the simulated admixture proportions α_{sim} or α sampled from a normal distribution $\mathcal{N}(\mu = \alpha_{sim}, \sigma^2 = 0.026^2 = 0.000676)$, 0.026 being the highest root-mean-square deviation of the ADMIXTURE estimation.[54]

(B) Receiver operating characteristic (ROC) curves comparing the performance of F_{adm}, with and without applying an allele frequency filter based on the source populations (see Material & Methods), under the 3 explored admixture with selection scenarios.

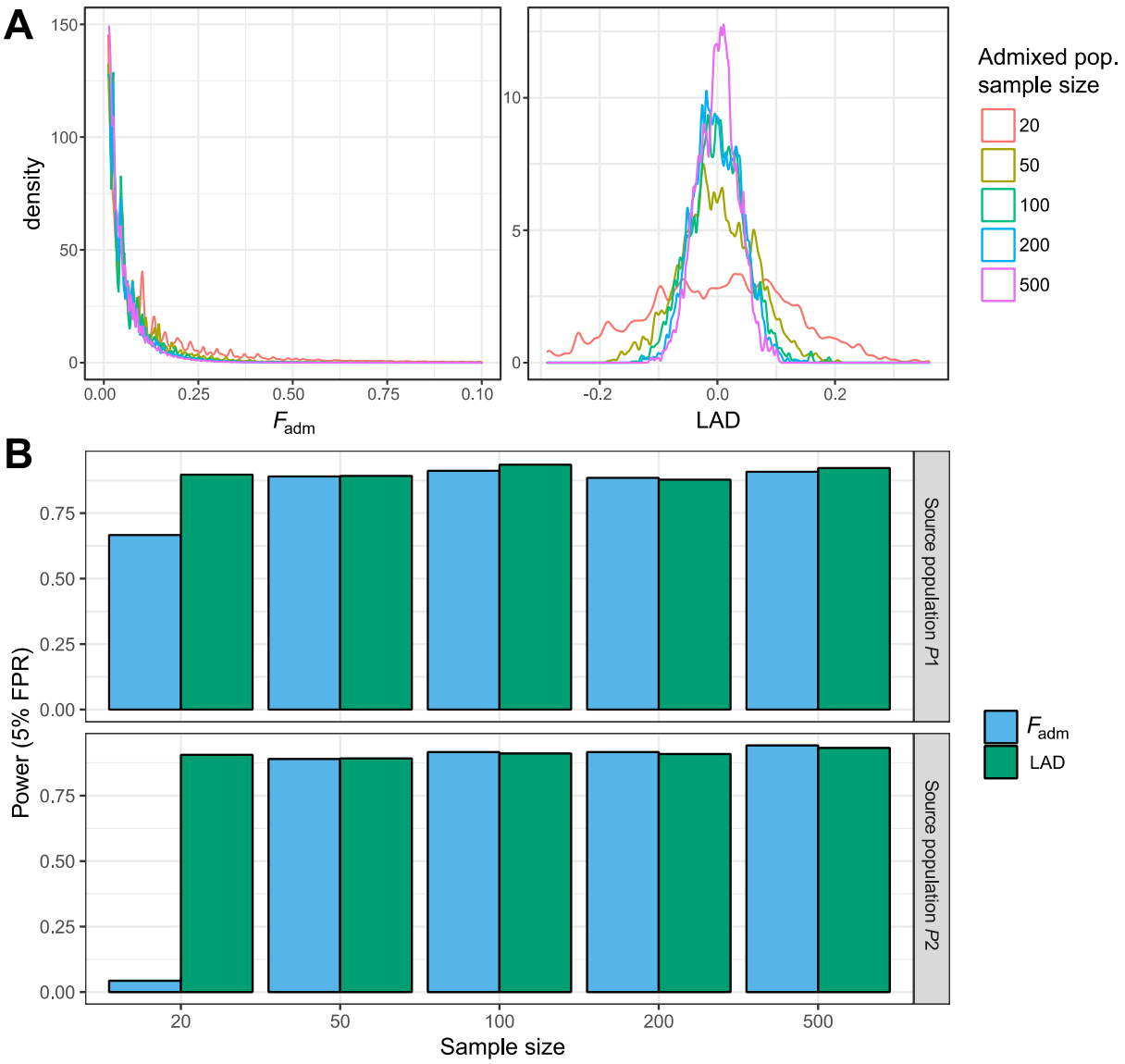

Figure S4. Effects of sample size on the power of F_{adm} and LAD statistics.

(A) Distributions under the null hypothesis (no positive selection) of F_{adm} and LAD, with varying sample sizes for the admixed population.

(B) Effect of the sample size of the source populations on the detection power of F_{adm} and LAD.

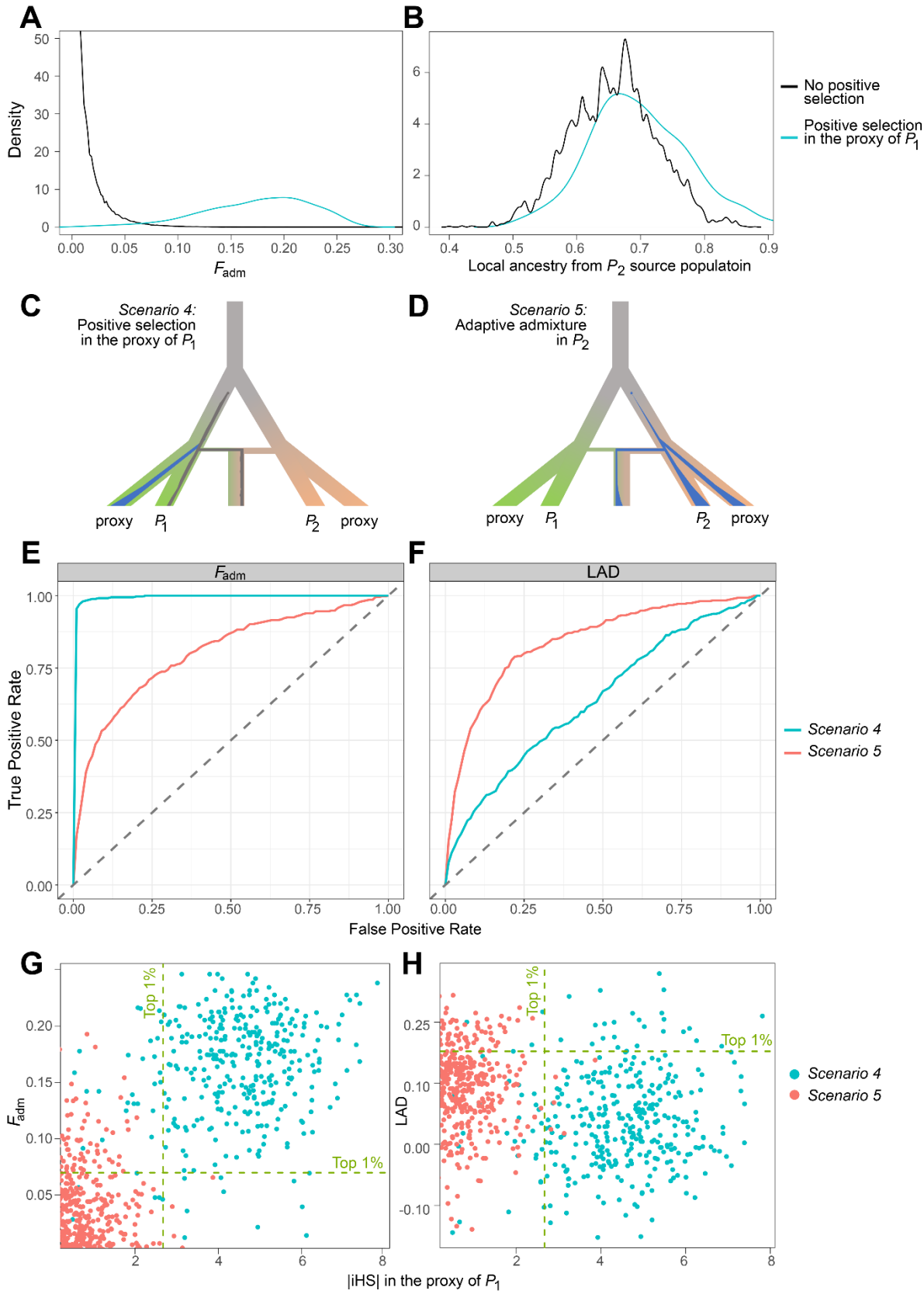

Figure S5. False positive signals due to selection in the proxy source population.

(A) Distributions of F_{adm} when there is or not positive selection in the proxy of the P_1 source population.

(B) Distributions of local ancestry in the admixed population from the P_2 source population, when there is or not positive selection in the proxy source population.
(C) The simulated model, assuming positive selection only in the proxy of the P_1 source population.
(D) The simulated model, assuming adaptive admixture in the P_2 source population. The scenario was simulated for comparison purposes.
(C-D) The blue and gray points indicate the appearance of a new beneficial and neutral mutations, respectively. The blue and gray areas indicate changes in frequency of the beneficial and neutral mutation, respectively.
(E-F) ROC curves for (E) F_{adm} and (F) LAD comparing the scenario where there is positive selection in the proxy of P_1 only (*scenario 4*; Figure S5C) and the scenario where there is a adaptive admixture in P_2 (*scenario 5*; Figure S5D).
(G–H) Absolute iHS values for the selected mutation in the proxy of the P_1 source population *vs.* (G) F_{adm} and (H) LAD values in the admixed population, when there is selection in this proxy of P_1 only (*scenario 4*; Figure S5C), or when there is adaptive admixture in P_2 (*scenario 5*; Figure S5D). Dashed green lines represent the 99^{th} percentiles (based on the null model simulations) for absolute iHS (vertical) and F_{adm} or LAD (horizontal). Excluding values that are above the absolute iHS 99^{th} percentile excludes approximately 90% of the extreme F_{adm} and LAD values under selection in this proxy of P_1 only (*scenario 4*) but, importantly, does not exclude any extreme value generated under the true adaptive admixture scenario (*scenario 5*).

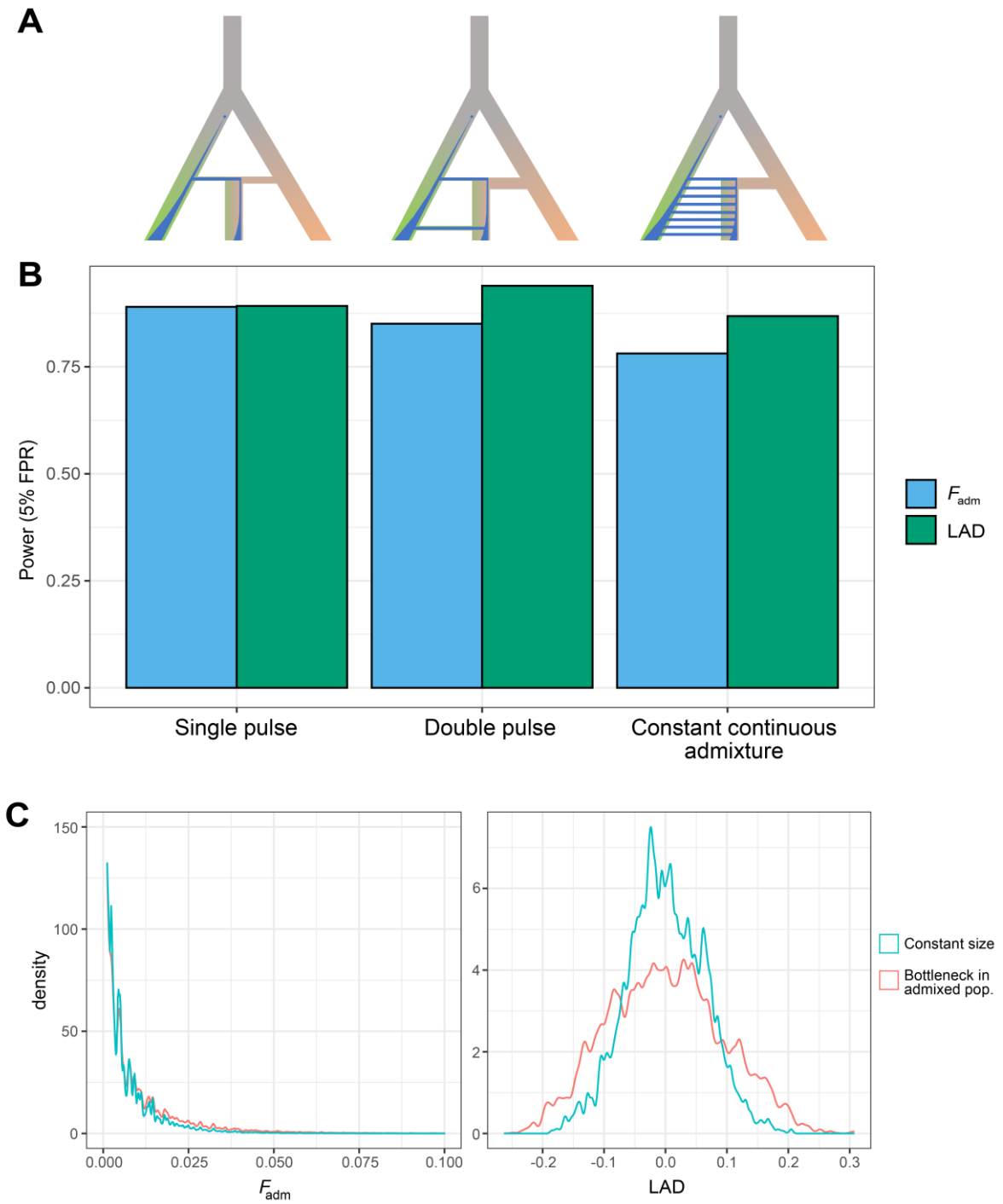

Figure S6. Effects of complex admixture and non-stationary demography on the power to detect adaptive admixture.

(A) The different simulated admixture models: a single pulse admixture model, a double pulse admixture model and a constant continuous admixture model. For these scenarios to be comparable, we set the sum of the admixture proportions contributed by each pulse to be equal to α_1 = 35%, and the average of the admixture dates to be equal to 70 generations (Material and Methods).

(B) Detection power of F_{adm} and LAD under the three different admixture scenarios (FPR = 5%; Material & Methods).

(C) Distributions of F_{adm} and LAD under the null hypothesis (no positive selection), with or without a 10-fold bottleneck in the admixed population.

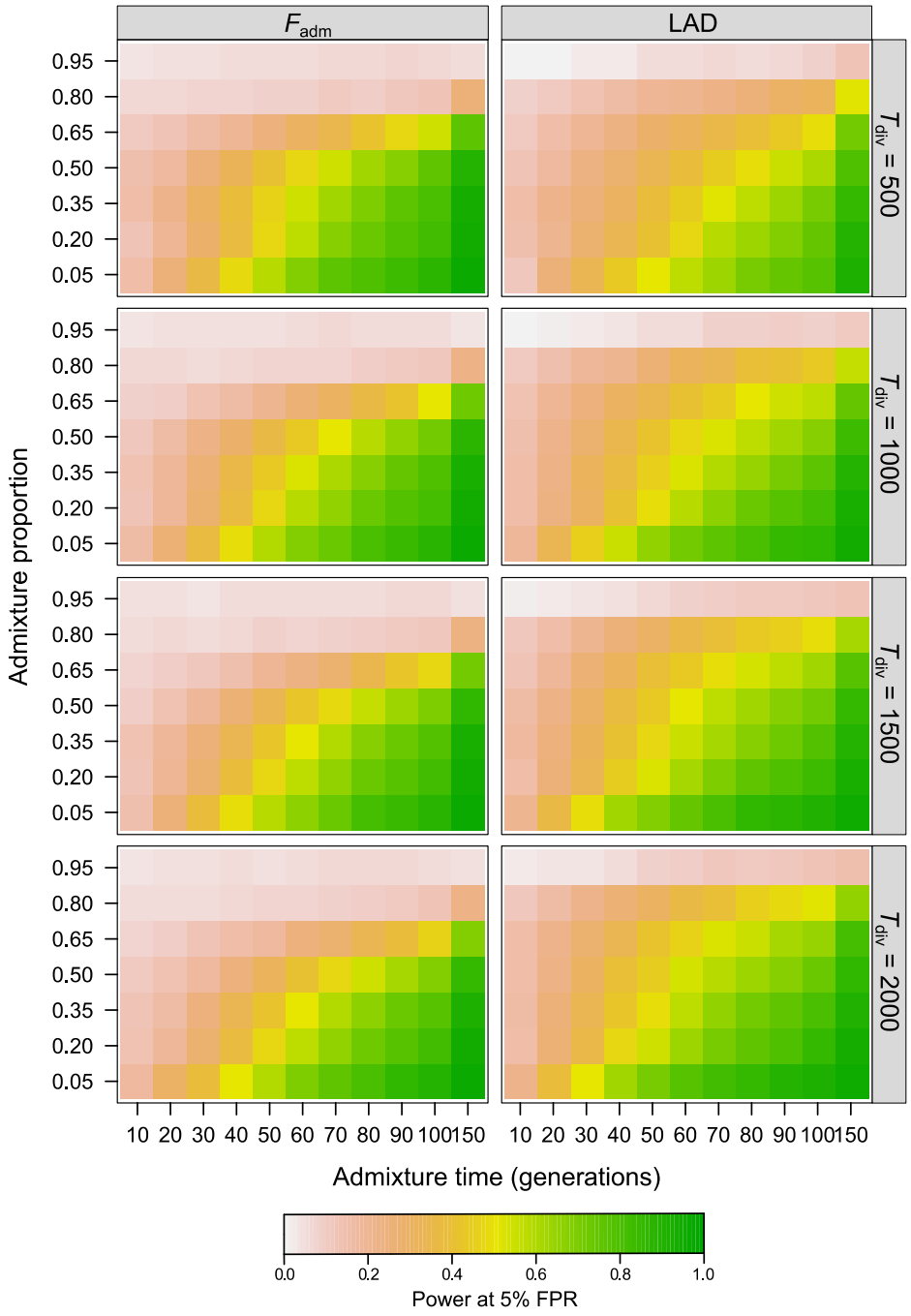

Figure S7. Effects of the divergence time between source populations on the power to detect adaptive admixture.

Effects on the detection power of F_{adm} and LAD of admixture time T_{adm}, admixture proportion α and the divergence time between source populations T_{div}. Colour indicates average detection power for a FPR = 5% threshold, across combinations of the remaining parameters. Because T_{div} is the upper limit of the time at which the beneficial mutation appears T_{mut}, we assumed for these simulations $T_{mut} < 500$ generations and $s \in \{0.05; 0.10\}$.

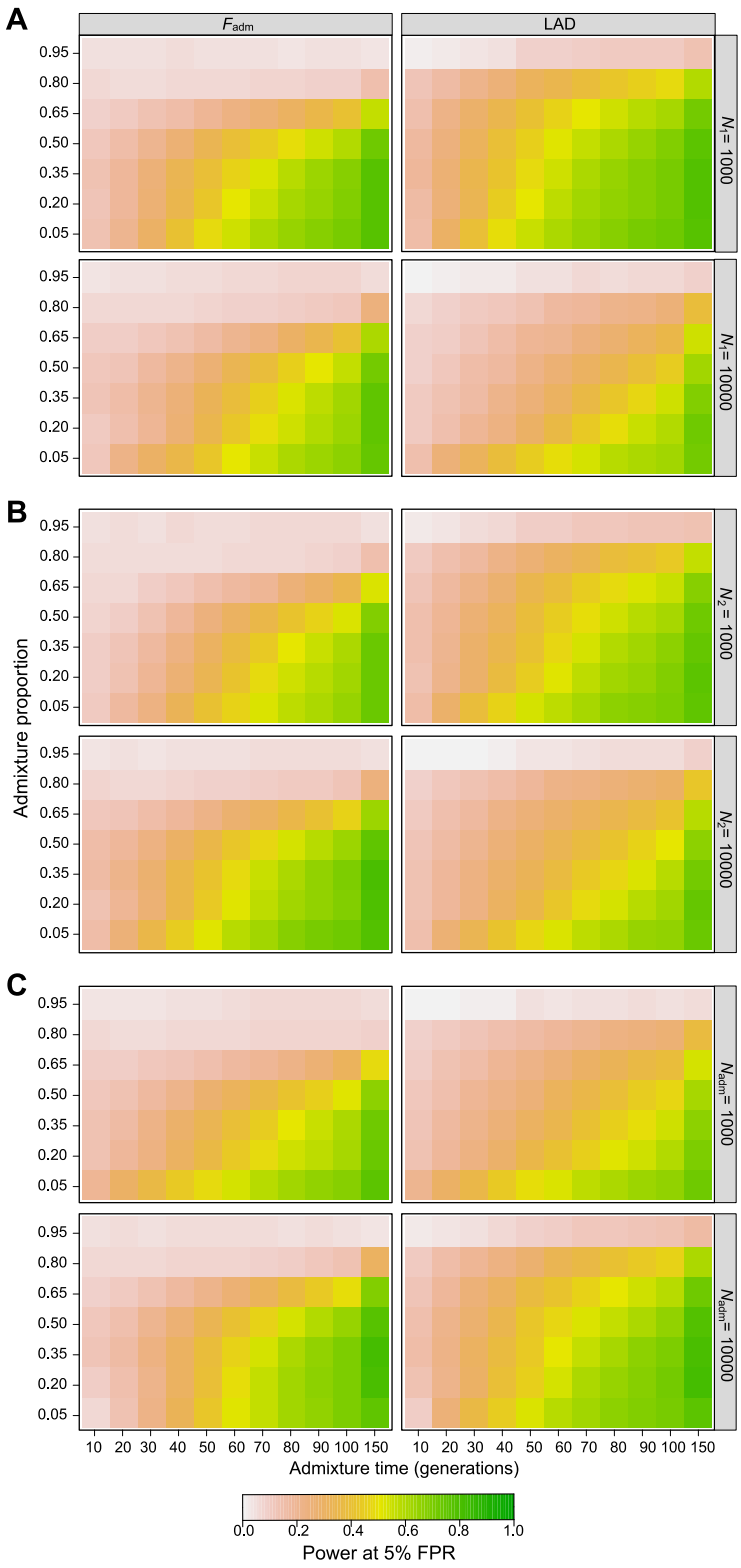

Figure S8. Effects of population sizes on the power to detect adaptive admixture.

Effects on the detection power of F_{adm} and LAD of admixture time T_{adm}, admixture proportion α and (A) N_1, (B) N_2 and (C) N_{adm}, the population sizes of source population P_1, source population P_2 and the admixed population, respectively (Figure S1). Colour indicates average detection power for a FPR = 5% threshold, across combinations of the remaining parameters.

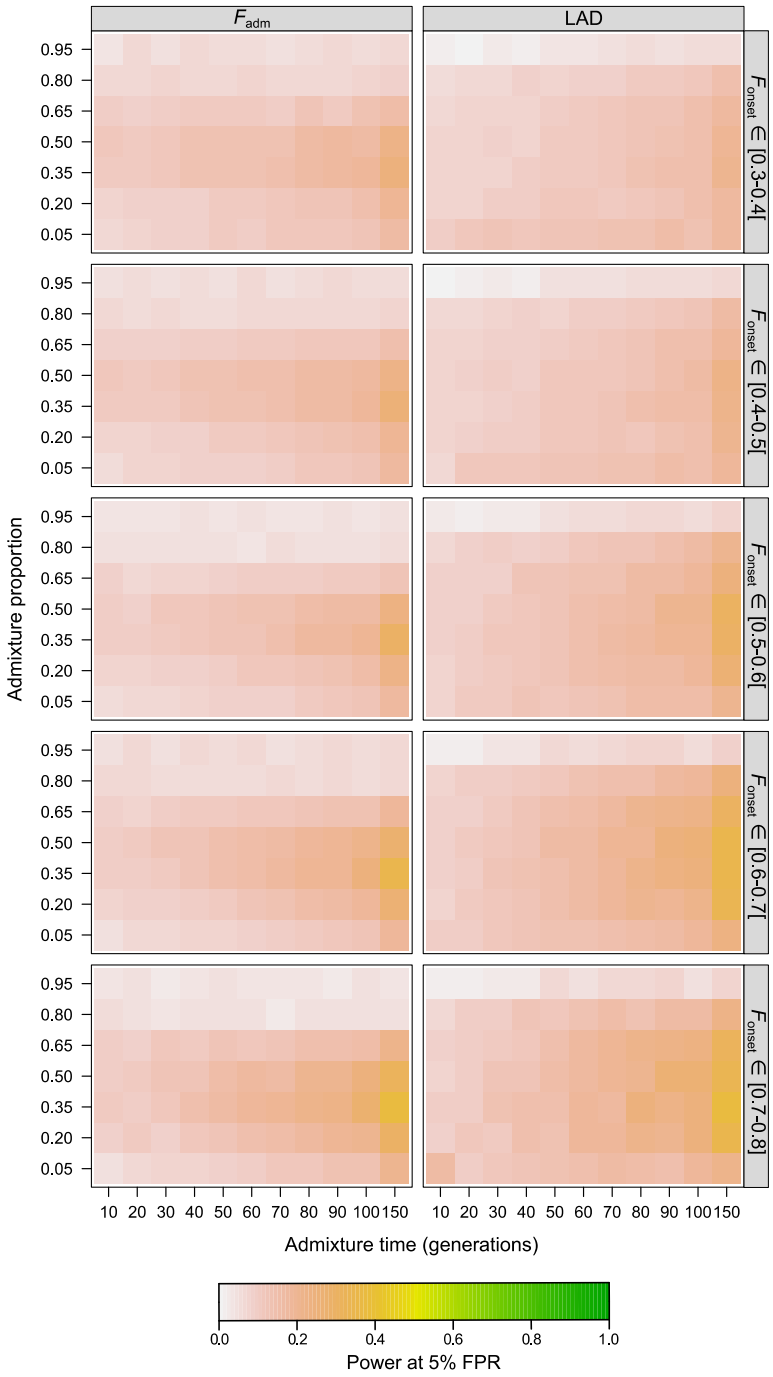

Figure S9. Effects of the frequency of the beneficial mutation (*s* = 0.01) on the power to detect adaptive admixture.

Effects on the detection power of F_{adm} and LAD of admixture time T_{adm}, admixture proportion α and F_{onset}, the frequency of the beneficial mutation in the source population at the time of admixture T_{adm}. Colour indicates average detection power for a FPR = 5% threshold, across combinations of the remaining parameters.

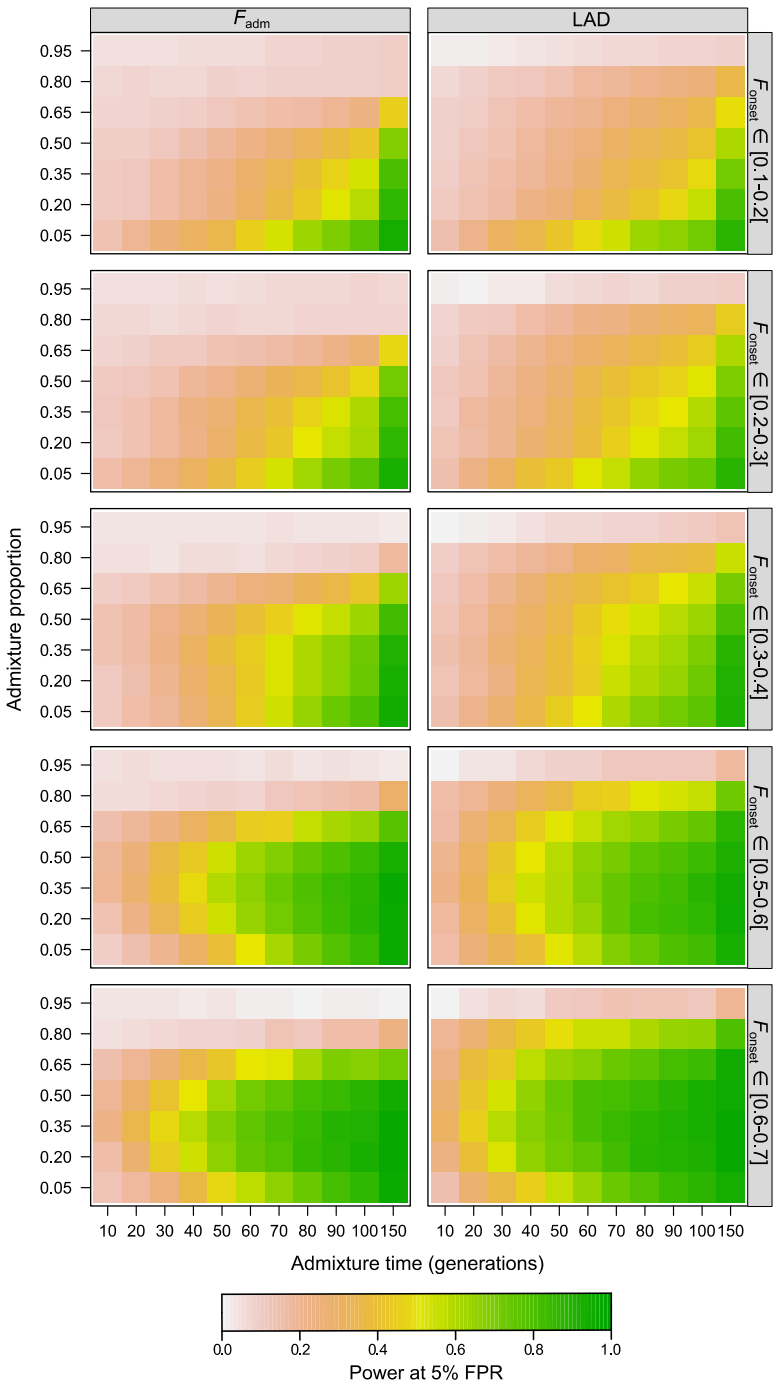

Figure S10. Effects of the frequency of the beneficial mutation (s = 0.05) on the power to detect adaptive admixture.

Effects on the detection power of F_{adm} and LAD of admixture time T_{adm}, admixture proportion α and F_{onset}, the frequency of the beneficial mutation in the source population at the time of admixture T_{adm}. Colour indicates average detection power for a FPR = 5% threshold, across combinations of the remaining parameters.

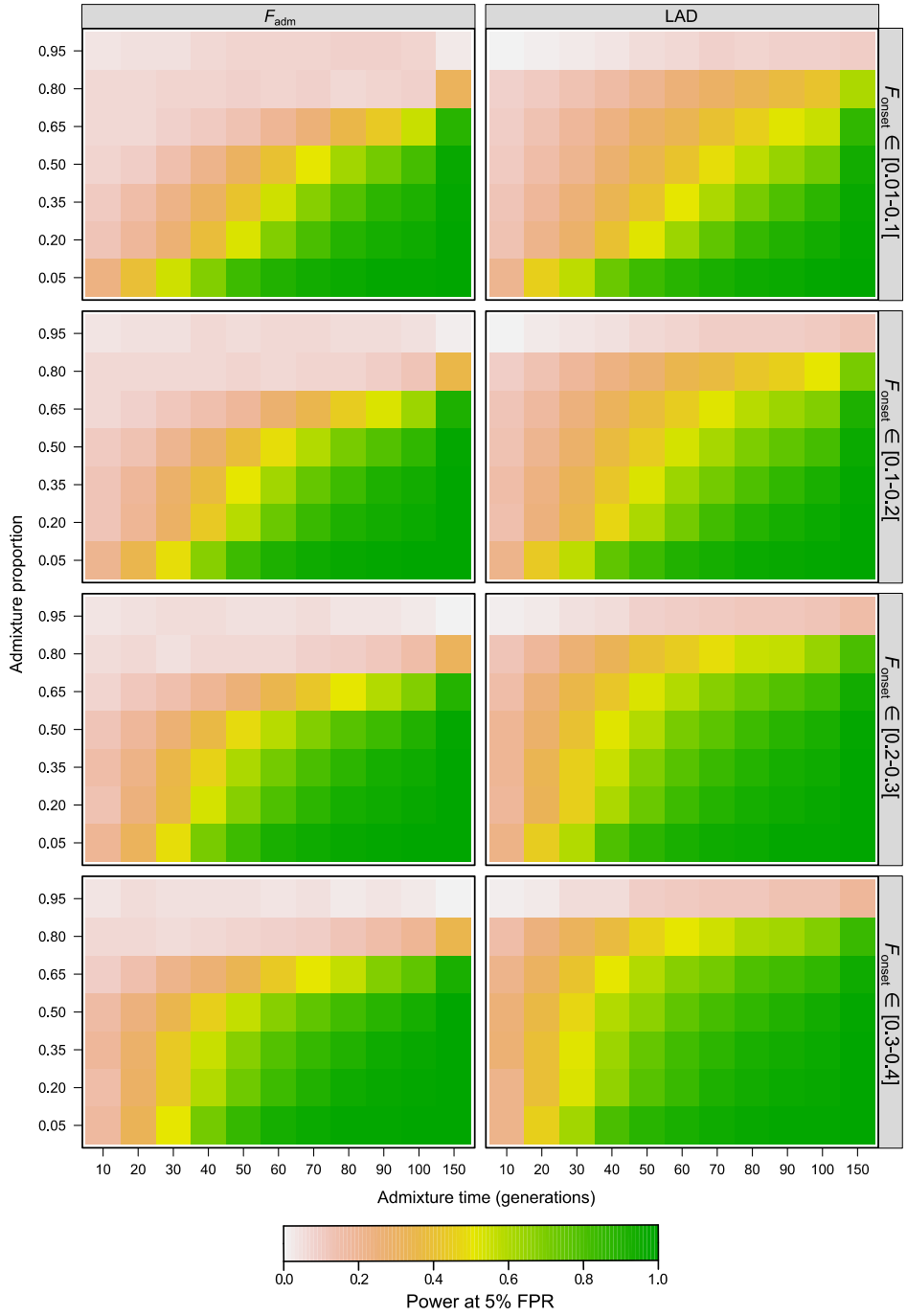

Figure S11. Effects of the frequency of the beneficial mutation (s = 0.10) on the power to detect adaptive admixture.

Effects on the detection power of F_{adm} and LAD of admixture time T_{adm}, admixture proportion α and F_{onset}, the frequency of the beneficial mutation in the source population at the time of admixture T_{adm}. Colour indicates average detection power for a FPR = 5% threshold, across combinations of the remaining parameters.

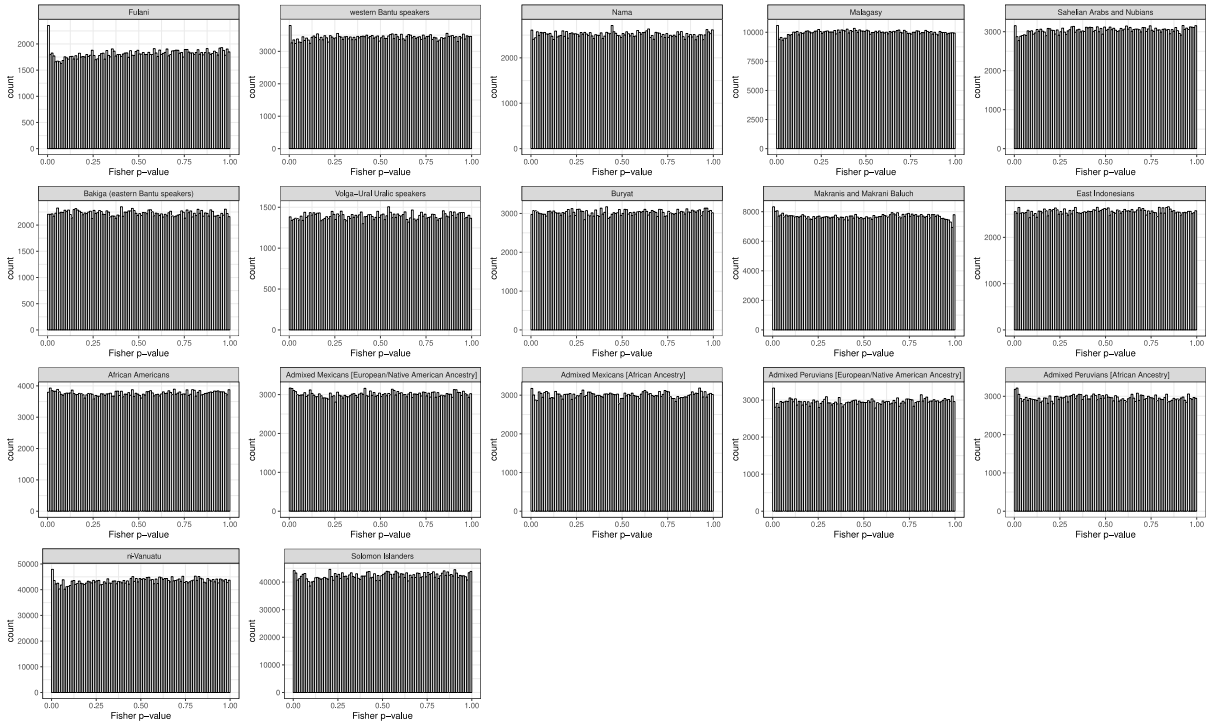

Figure S12. Distributions of Fisher's combined *P*-values in the empirical data. Histograms of combined *P*-values using Fisher's method, for the 15 analysed admixed populations. The *P*-values are uniformly distributed, except for certain populations where there is an excess of small *P*-values, corresponding to the populations where signals for adaptive admixture were found.

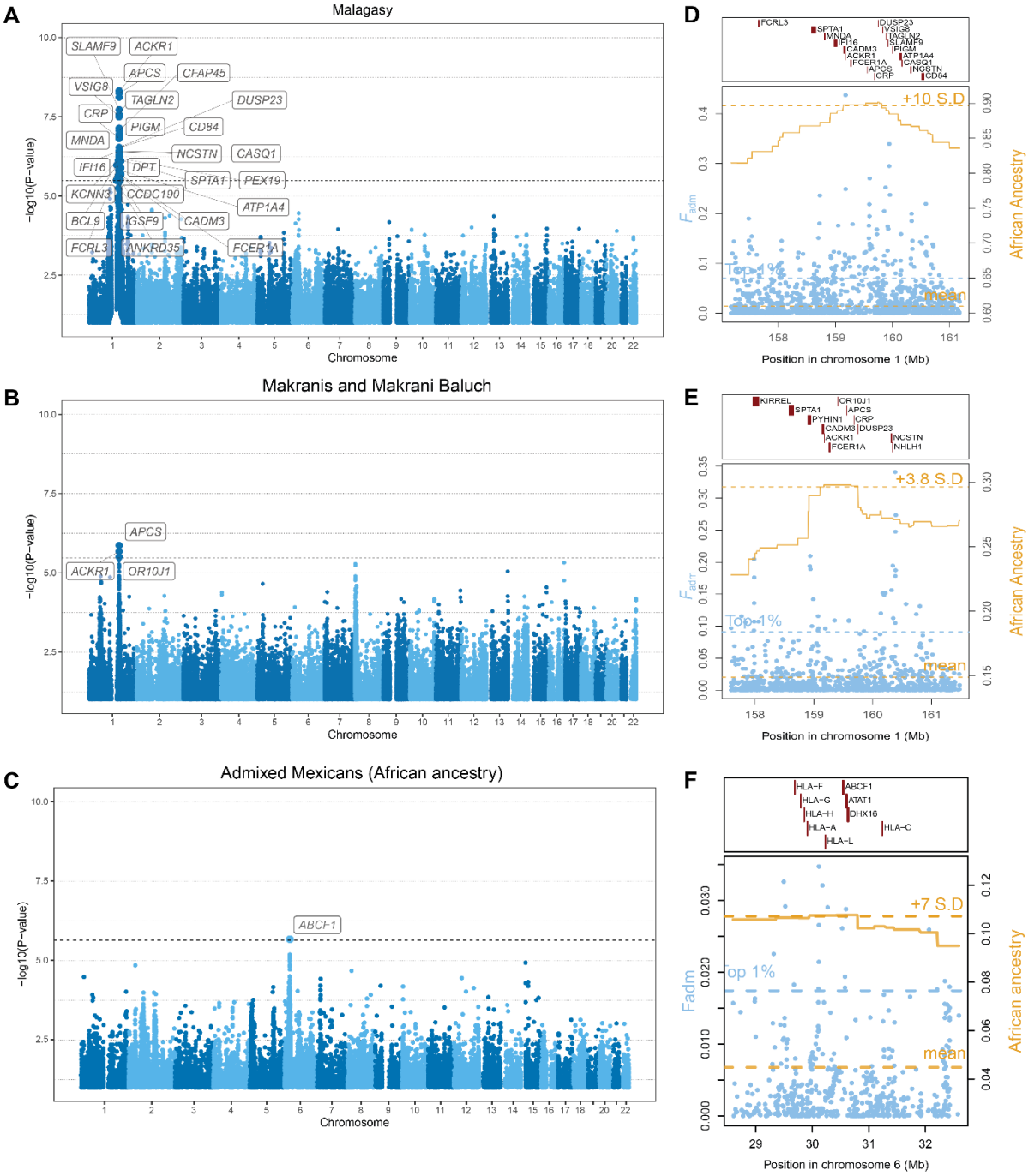

Figure S13. Other previously reported genomic signals of adaptive admixture.

(A) Genome-wide signals of adaptive admixture in Malagasy populations from Madagascar.

(B) Genome-wide signals of adaptive admixture in African-descent Makranis and Makrani Baluch from Pakistan.

(C) Genome-wide signals of adaptive admixture in admixed Mexicans (African ancestry).

(A-C) Highlighted blue points indicate variants that passed the Bonferroni significance threshold (shown by a horizontal dotted line). Gene labels were attributed based on the gene with the highest V2G score within 250-kb of the candidate variant.

(D) Local signatures of adaptive admixture for the *ACKR1* region in Malagasy from Madagascar.
(E) Local signatures of adaptive admixture for the *ACKR1* region in Makranis and Makrani Baluch from Pakistan.
(F) Local signatures of adaptive admixture for the *HLA* class I region in admixed Mexicans.
(D-F) Light blue points indicate F_{adm} values for individual variants. The gold solid line indicates the average African local ancestry.

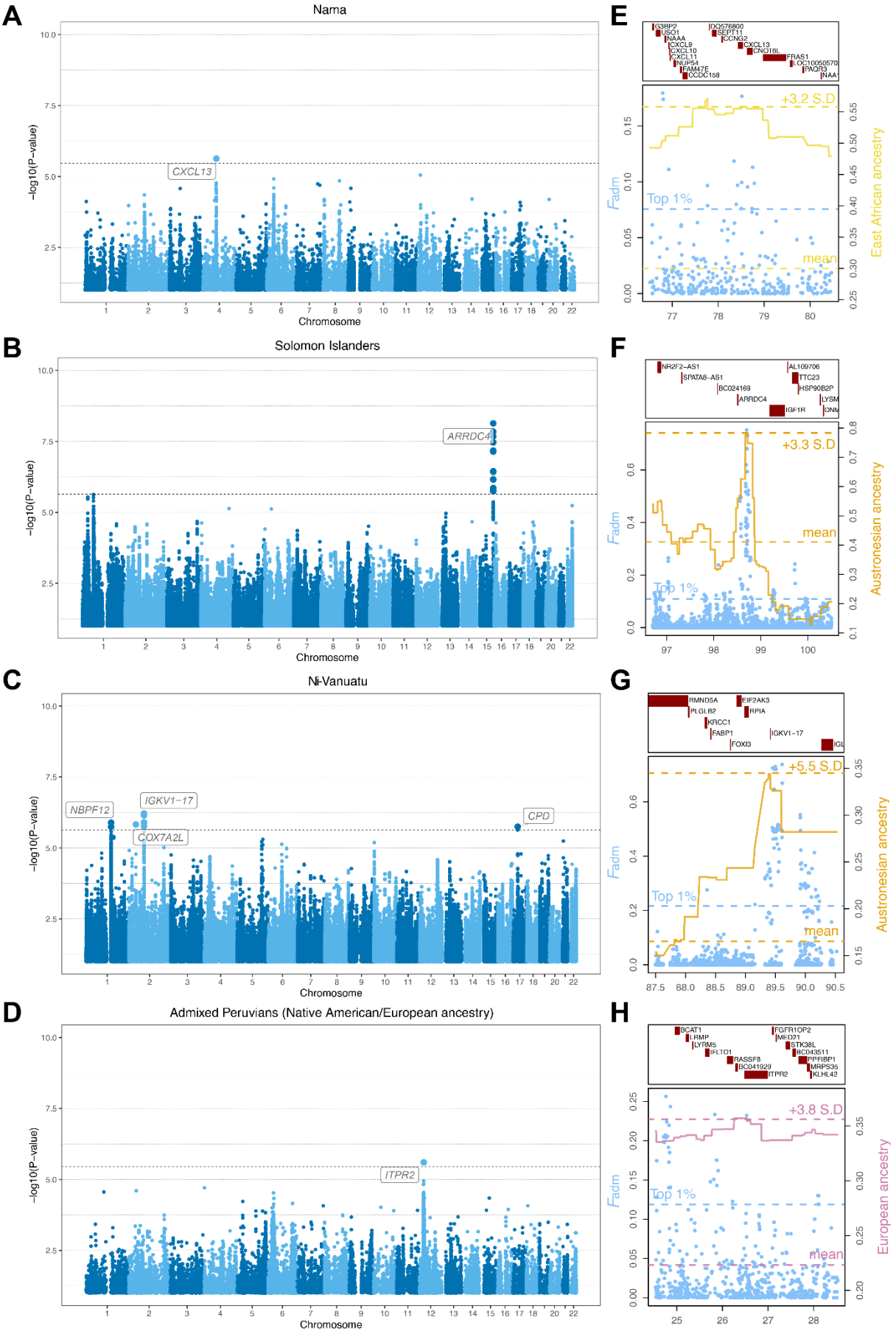
A
Nama
−log10(P−value)
CXCL13
Chromosome
E
+3.2 S.D
Top 1%
mean
Fadm
East African ancestry
B
Solomon Islanders
ARRDC4
F
+3.3 S.D
Austronesian ancestry
C
Ni-Vanuatu
NBPF12
IGKV1−17
COX7A2L
CPD
G
+5.5 S.D
D
Admixed Peruvians (Native American/European ancestry)
ITPR2
H
+3.8 S.D
European ancestry

Figure S14. Other novel genomic signals of adaptive admixture.

(A) Genome-wide signals of adaptive admixture in the Nama from South Africa.

(B) Genome-wide signals of adaptive admixture in Solomon Islanders.

(C) Genome-wide signals of adaptive admixture in Vanuatu Islanders.

(D) Genome-wide signals of adaptive admixture in admixed Peruvians.

(A-D) Highlighted blue points indicate variants that passed the Bonferroni significance threshold (shown by a horizontal dotted line). Gene labels were attributed based on the gene with the highest V2G score within 250-kb of the candidate variant.

(E) Local signatures of adaptive admixture for the *CNOT6L/CXCL13* region in the Nama from South Africa.

(F) Local signatures of adaptive admixture for the *ARRDC4* region in Solomon Islanders.

(G) Local signatures of adaptive admixture for the *IGKV1-17* region in Vanuatu Islanders.

(H) Local signatures of adaptive admixture for the *ITPR2* region in admixed Peruvians.

(E-H) Light blue points indicate F_{adm} values for individual variants. The yellow, gold and pink solid lines indicate average local ancestry from East Africans, Austronesians and Europeans respectively.

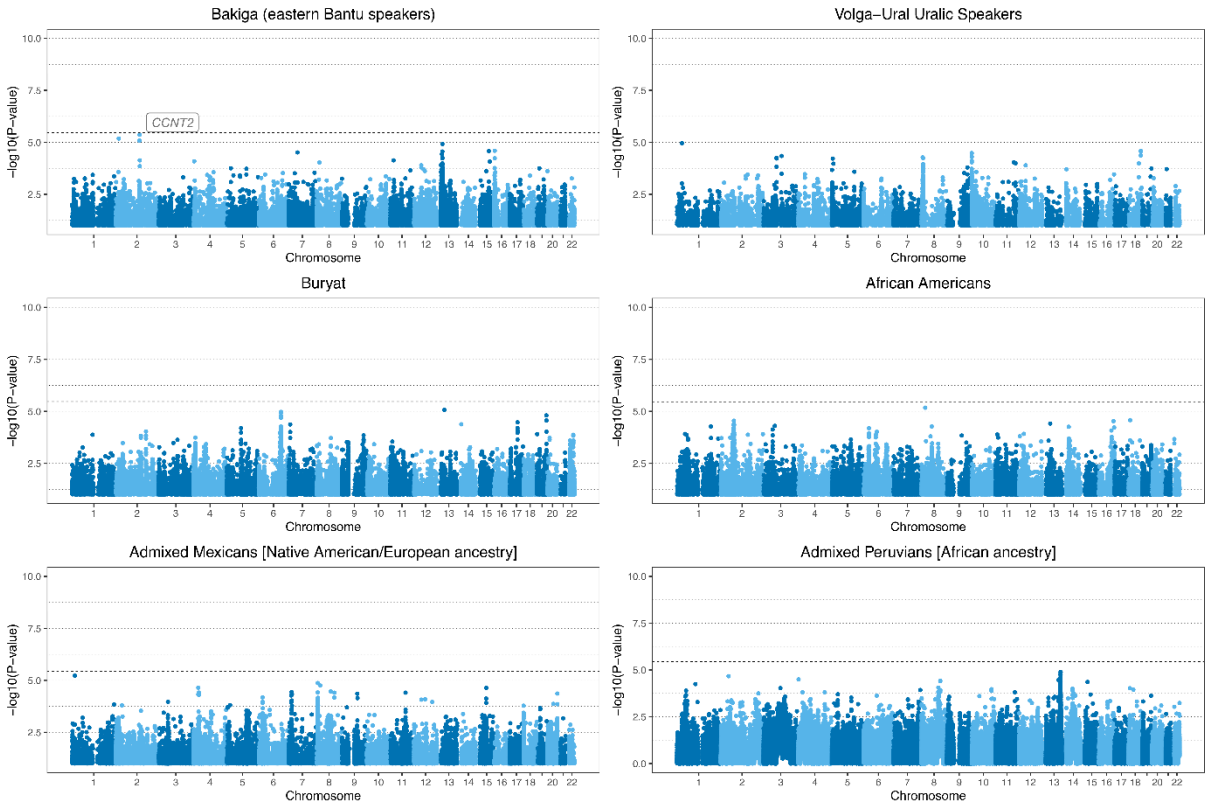

Figure S15. Genome scans for populations where there is no evidence for adaptive admixture.

Manhattan plots of $-\log_{10}(P\text{-values})$ for the combined Fisher's method, in the remaining 6 admixed populations where no variant passes the Bonferroni significance threshold (shown by a horizontal dotted line).