The American Journal of Human Genetics, Volume *103*

Supplemental Data

Natural Selection Has Differentiated the Progesterone Receptor among Human Populations

Jingjing Li, Xiumei Hong, Sam Mesiano, Louis J. Muglia, Xiaobin Wang, Michael Snyder, David K. Stevenson, and Gary M. Shaw

SUPPLEMENTAL DATA

Supplemental data includes 8 figures and 4 tables

Figure.S1 XP-EHH scan to detect positive selection in the *PGR* locus. The peaks indicate signatures of positive selection leading to extreme extended haplotype homozygosity in one population but not the other. Comparisons were performed for CHB-CEU, CHB-YRI and CEU-YRI. Significance was empirically determined by the upper 5% XP-EHH scores across the entire genome for each comparison (the grey horizontal line). XL-EHH scores were computed at each individual SNP locus.

Figure.S2 Nucleotide diversity in the *PGR* locus. The *PGR* locus displayed increased nucleotide diversity in CEU, and reduced diversity in CHB. For each comparison, nucleotide diversity was compared between the *PGR* locus and the genome background.

Figure.S3 EHH test on the *PGR* locus. The EHH (extended haplotype homozygosity) test was performed on the *PGR* locus in CHB, CEU and YRI, where significant signals were only identified in CHB. Statistical significance was determined by the empirical P values across the genome in each population indicated by the dotted blue lines.

Figure.S4 Gene expression of *PGR* across human tissue and cell types, determined by RNA-Seq from the GTEx Project.

Figure.S5. A significant correlation between the effect size on PGR ovary expression and and population differentiation (*Fst* between CHB and CEU) for the HD-SNPs. The SNP rs11224580 represents an exemplar site with a strong effect on modulating *PGR* in the ovary, and its high derived allele frequency is specific in CHB, but not in CEU and YRI.

Figure.S6 Extreme haploinsufficiency of *PGR.* The haploinsufficiency score of *PGR* accounts for the upper 3% across all the human genes. The predicted haploinsufficiency scores were from a previous study by Huang *et al.*

Figure.S7 Distribution of *Fst* for the HD-SNPs. The population differentiation, *Fst*, was computed between CHB and YRI. The bimodal distribution is evident, indicating the enrichment of both high-*Fst* (*Fst*≥0.5) and low *Fst* (*Fst*≤0.1) sites.

Figure.S8 ATAC-Seq peaks in the ovary. ATAC-Seq peaks indicate the location of potential regulatory elements in the ovary. ATAC-Seq peaks with intensity fold change greater than 2 over control was considered significant.

Supplemental Tables (in separate spreadsheets)

Table S1. The *PGR* locus with high derived allele frequencies (≥0.7) in CHB.

Table S2. The common SNPs for linkage disequilibrium analysis.

Table S3. The effect size of HD-SNPs on modulating PGR expression in the breast mammary tissue, uterus, vagina and ovary.

Supplemental Figures

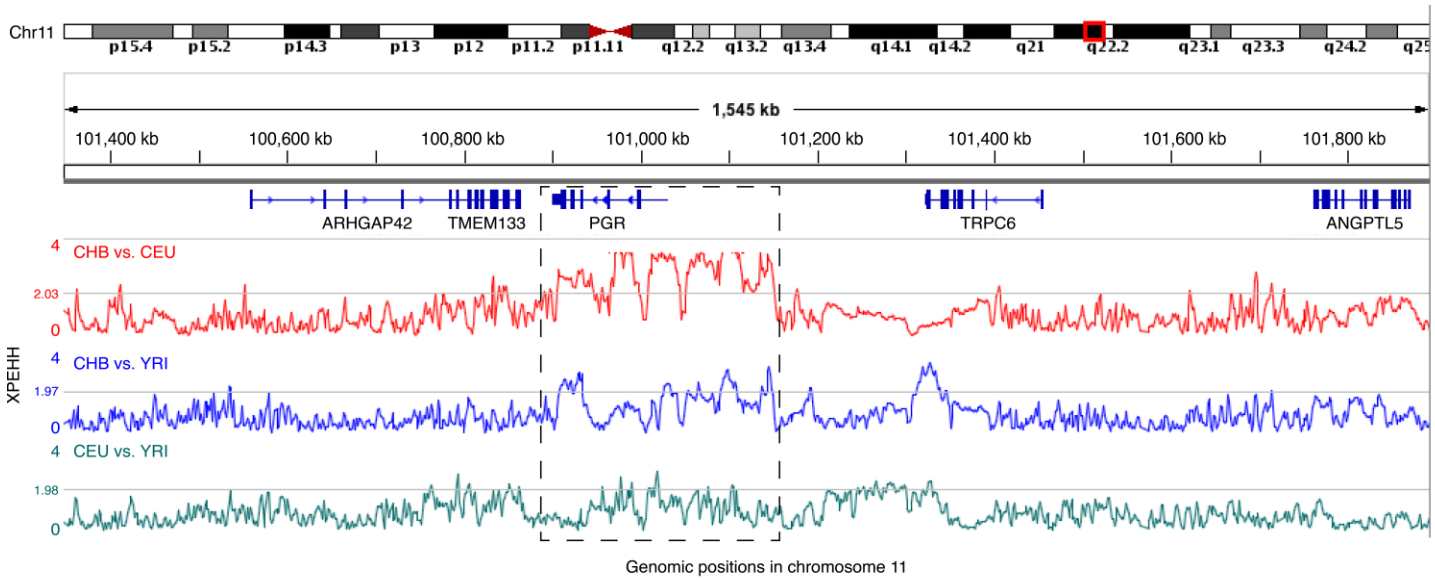

Figure.S1 XP-EHH scan to detect positive selection in the *PGR* locus. The peaks indicate signatures of positive selection leading to extreme extended haplotype homozygosity in one population but not the other. Comparisons were performed for CHB-CEU, CHB-YRI and CEU-YRI. Significance was empirically determined by the upper 5% XP-EHH scores across the entire genome for each comparison (the grey horizontal line). XL-EHH scores were computed at each individual SNP locus.

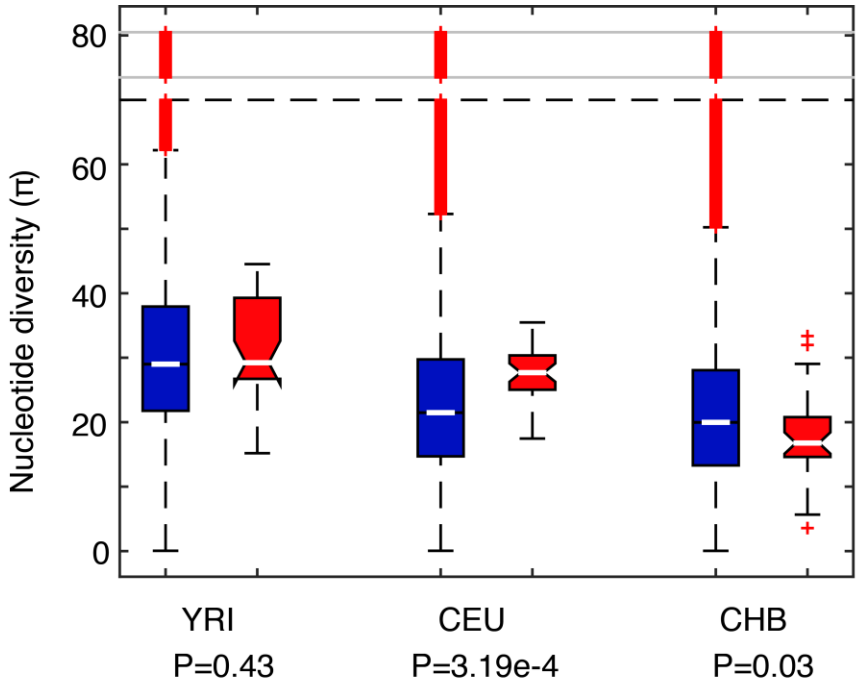

Figure.S2 Nucleotide diversity in the *PGR* locus. The *PGR* locus displayed increased nucleotide diversity in CEU, and reduced diversity in CHB. For each comparison, nucleotide diversity was compared between the *PGR* locus and the genome background.

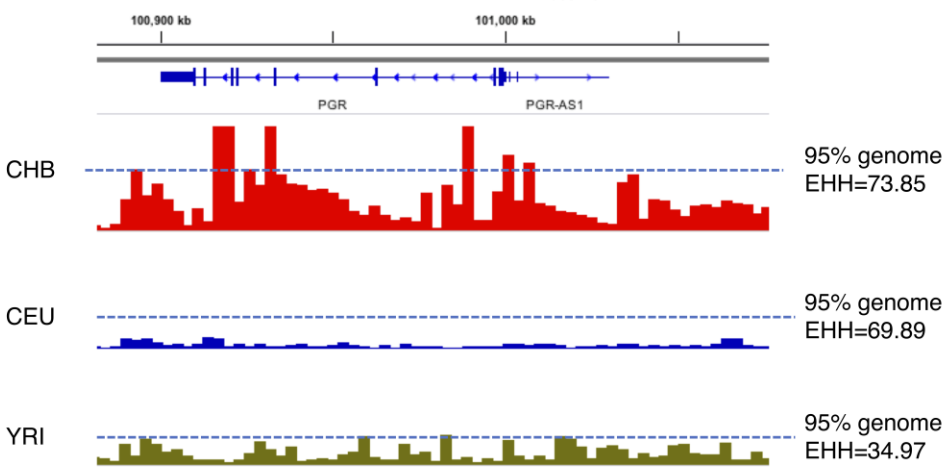

Figure.S3 EHH test on the *PGR* locus. The EHH (extended haplotype homozygosity) test was performed on the *PGR* locus in CHB, CEU and YRI, where significant signals were only identified in CHB. Statistical significance was determined by the empirical P values across the genome in each population indicated by the dotted blue lines.

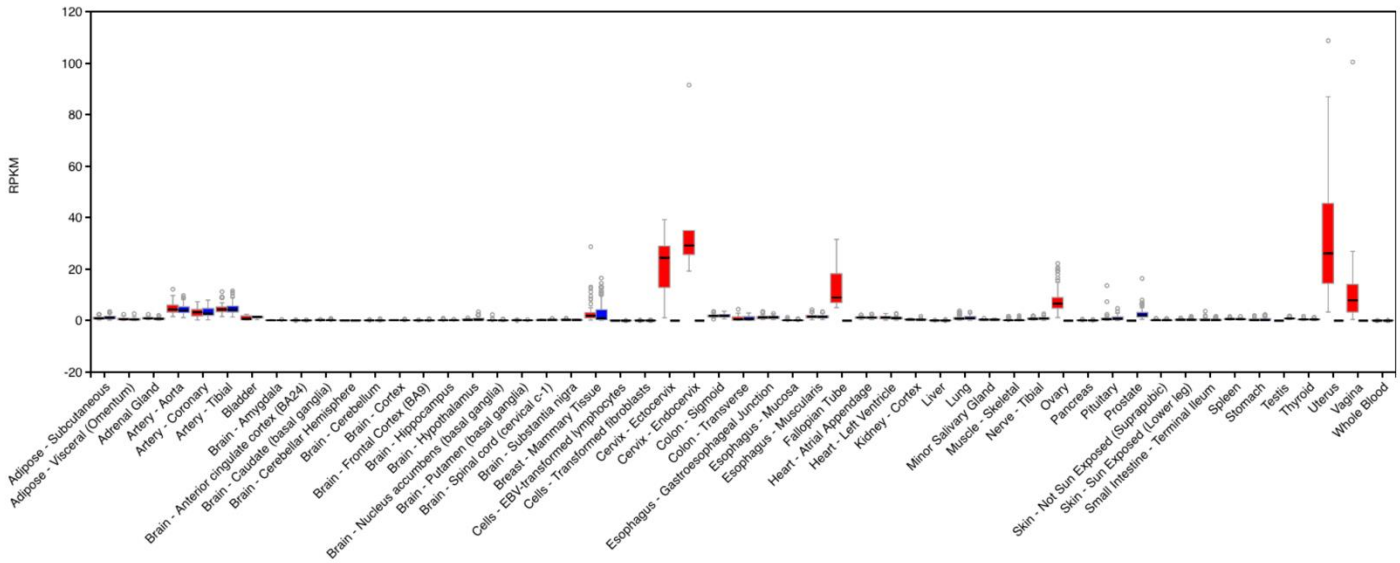

Figure.S4 Gene expression of *PGR* across human tissue and cell types, determined by RNA-Seq from the GTEx Project.

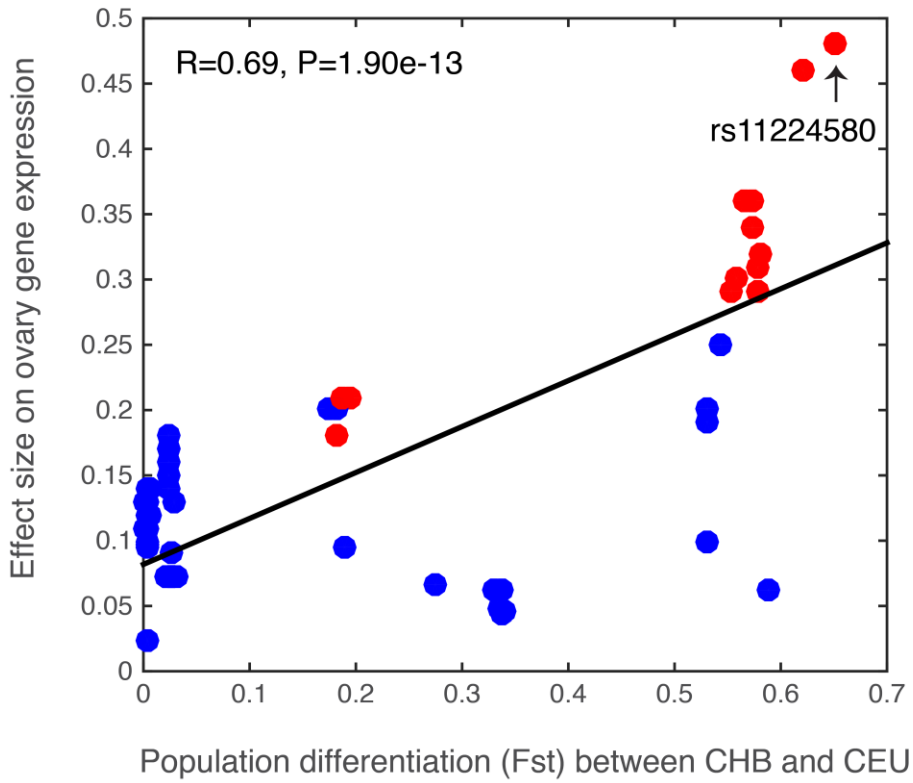

Figure.S5. A significant correlation between the effect size on PGR ovary expression and and population differentiation (*Fst* between CHB and CEU) for the HD-SNPs. The SNP rs11224580 represents an exemplar site with a strong effect on modulating *PGR* in the ovary, and its high derived allele frequency is specific in CHB, but not in CEU and YRI.

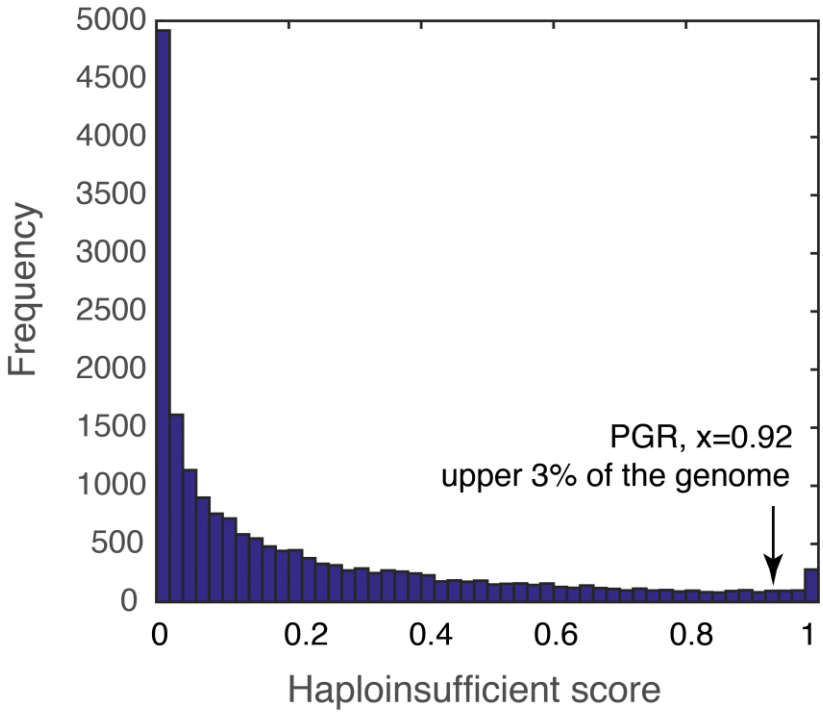

Figure.S6 Extreme haploinsufficiency of *PGR*. The haploinsufficiency score of *PGR* accounts for the upper 3% across all the human genes. The predicted haploinsufficiency scores were from a previous study by Huang *et al.*

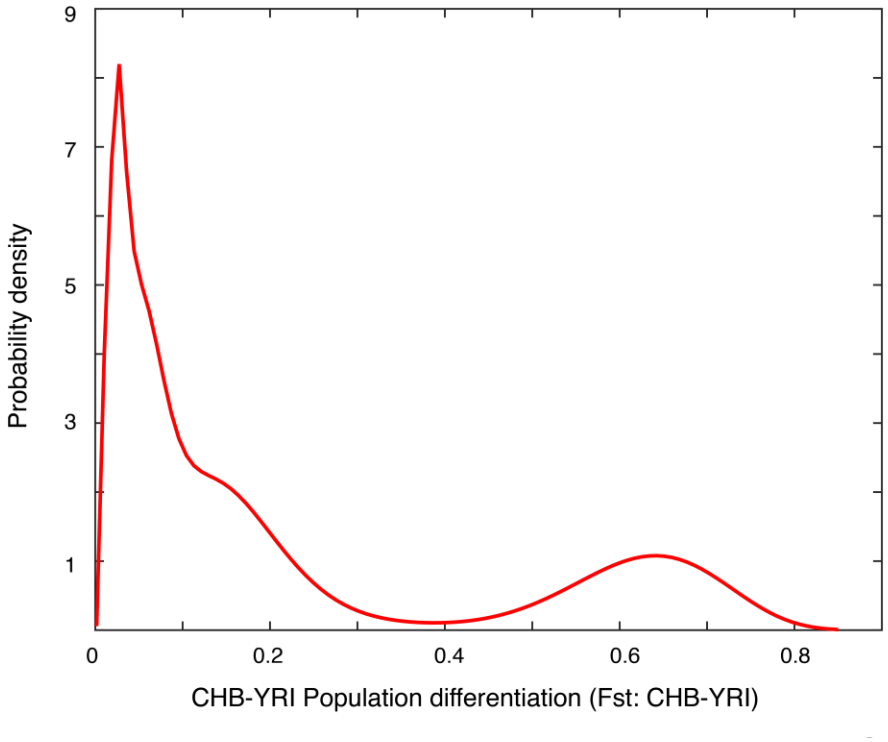

Figure.S7 Distribution of *Fst* for the HD-SNPs. The population differentiation, *Fst*, was computed between CHB and YRI. The bimodal distribution is evident, indicating the enrichment of both high-*Fst* (*Fst*≥0.5) and low *Fst* (*Fst*≤0.1) sites.

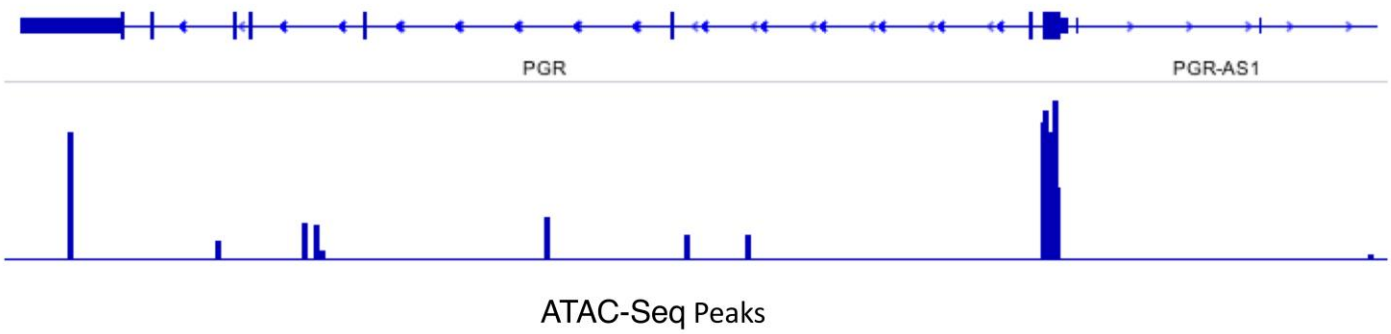

Figure.S8 ATAC-Seq peaks in the ovary. ATAC-Seq peaks indicate the location of potential regulatory elements in the ovary. ATAC-Seq peaks with intensity fold change greater than 2 over control was considered significant.