PCA for eMERGE study samples (n=17,150)

Legend:
- American Indian / Alaska Native (n=31)
- Asian (n=84)
- Black or African American (n=1,699)
- Other (n=61)
- White (n=12,170)
- CEU (n=34)
- YRI (n=20)
- CHB (n=3)
- JPT (n=2)
- Unknown (n=3,046; data not shown)
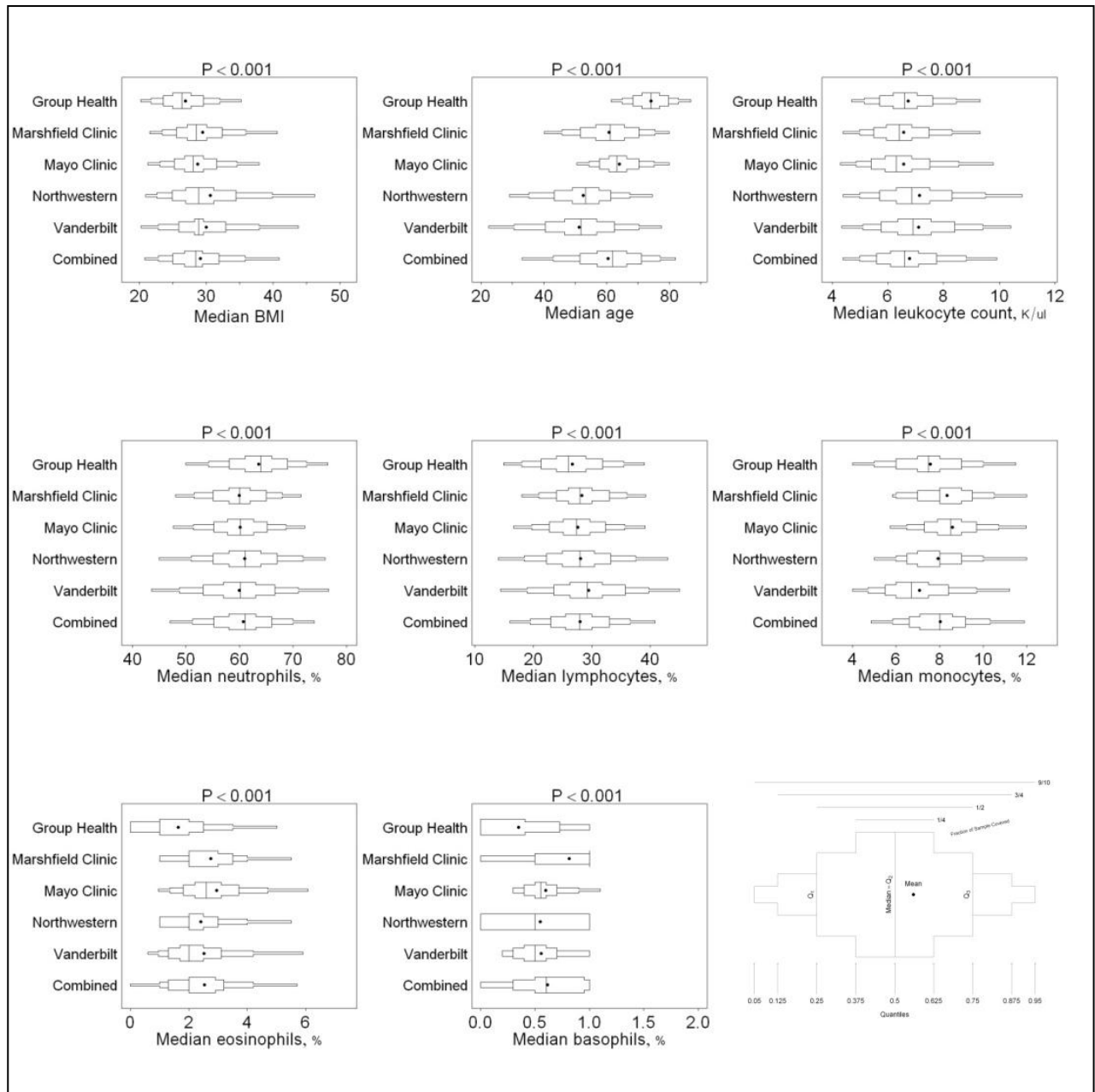
Eigenvector 2 (0.3%)

Eigenvector 1 (4.8%)

Supplemental Figure S1: Plots of eigenvectors 1 and 2 using PCA. Principal components analysis of 17,150 eMERGE subjects from all field centers, anchored by HapMap controls. Color-coding is according to self-identified or self-observed race.
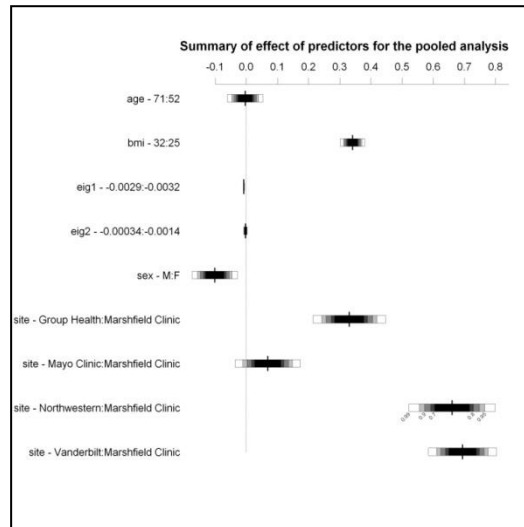
Supplemental Figure S2: Box-percentile plots of median age, median BMI, median leukocyte count and median differentials. The bottom right plot is an annotated example showing the layout of the box-percentile plots. P-values listed are from the Kruskal-Wallis test as listed in Table 2.
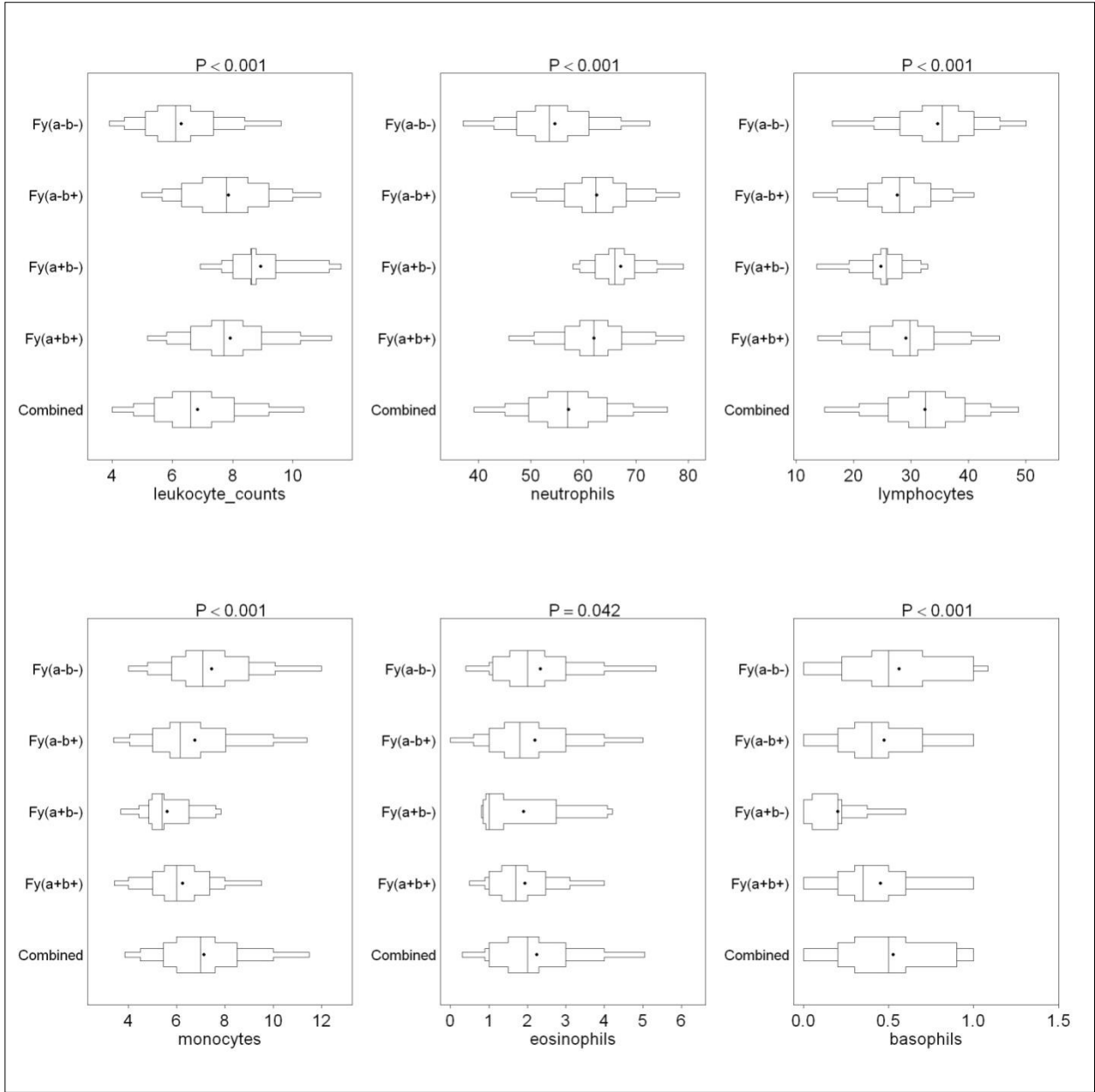
a.

b.

c.
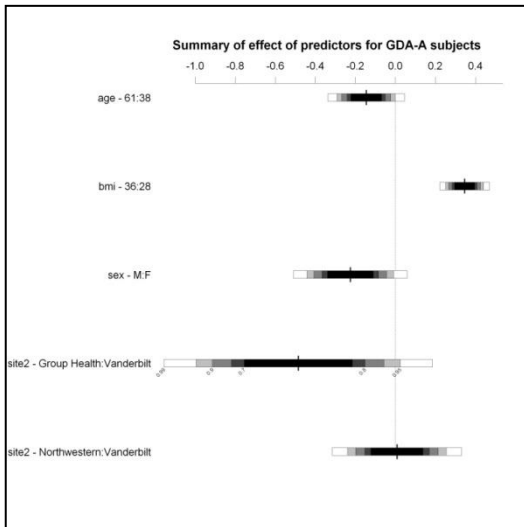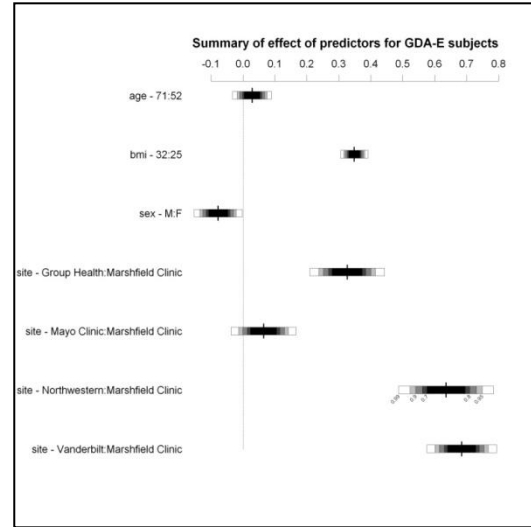
Figure S3: Summary of effects of predictors on WBC in the (a.) AA, (b.) EE and (c.) pooled

models. Continuous variables are compared using default ranges (interquartile).

Figure S4: Box-percentile plots of median leukocyte count and median differentials by Duffy phenotype (Fya+b+, Fya+b-, Fya-b+, Fya-b-) in the subjects of African ancestry. The bottom right plot in Figure 1 is an annotated example showing the layout of the box-percentile plots. P-values listed are from the Kruskal-Wallis test.

	Group Health		Marshfield		Mayo		Northwestern		Vanderbilt	
	ß	p-value	ß	p-value	ß	p-value	ß	p-value	ß	p-value
1q21, *DARC*										
rs12075	0.02	6.30E-01	-0.06	8.37E-02	-0.07	9.42E-02	0.10	2.27E-01	0.10	5.69E-02
17q21.1, *GSDMA*										
rs3894194	0.04	3.50E-01	0.09	5.06E-03	0.11	1.06E-02	0.19	1.27E-02	0.12	7.70E-03
rs3859192	0.09	2.48E-02	0.10	3.36E-03	0.20	3.37E-06	0.18	1.68E-02	0.13	4.06E-03
17q21.1, *PSMD3*										
rs4065321	0.14	5.27E-04	0.10	1.64E-03	0.15	7.60E-04	0.19	1.81E-02	0.15	3.95E-04
17q21.1, *MED24*										
rs9916158	-0.11	9.49E-03	-0.13	8.41E-05	-0.18	4.58E-05	-0.13	1.29E-01	-0.09	6.79E-02