Supplemental Information:

***Methods**:*

GABC Sample Collection, Genotyping, and Data Cleaning

50ml of blood was collected by venipuncture into 10% acid-citrate-dextrose (ACD) and centrifuged at 2000 x G for 10 minutes to obtain platelet-poor plasma and buffy coat, which were separated, aliquoted into coded de-identified cryovials and stored in liquid nitrogen. Thawed plasma samples were subsequently aliquoted into 96-well storage plates. Each plate contained 2 control wells carrying the laboratory standard plasma (FACT, George King Bio-medical, Overland Park, KS). Genomic DNA samples from 1,189 study subjects were extracted from buffy coat using the QIAcube (Qiagen, Valencia, CA) method, quantified, and genotyped on the Illumina Omni1-Quad v1 B array at the Broad Institute of MIT and Harvard (Cambridge, MA). Details of the extensive quality control process have been described(1) and additional details are described in a data cleaning report available at the Geneva study website.(2) The final "cleaned" dataset, for 763,195 SNPs and 1,152 subjects representing 489 sibships, was used in further analyses. GABC phenotype and genotype data have been posted to dbGaP.(3)

TSS Genotyping and Data Cleaning

Each participant donated 30ml of blood after completing a questionnaire regarding diet and dietary supplements related to folate metabolism. All samples

and linked data were de-identified prior to use. Investigators at the University of Michigan received 200ul of platelet-poor plasma that was anti-coagulated in EDTA and frozen in 96-well storage plates. The plates contained intra- and inter-plate duplicates and blank wells for adding laboratory standard plasma (FACT). Genomic DNA was extracted using the Qiagen QIAamp DNA Blood Mini Kit or the Qiagen DNeasy Kit (Qiagen, UK). SNP genotyping was conducted at the Center for Inherited Disease Research (CIDR) (Baltimore, Maryland), using Illumina 1M HumanOmni1-Quad_v1-0_B chips. Genotypes were attempted for 2,438 study samples, 14 blind duplicates, and 105 HapMap controls. The blind duplicates had a concordance rate of 99.997%; and the HapMap samples had a 99.71% concordance rate with their known genotypes.

Samples were excluded based on 1) incomplete phenotype information (n=11), 2) gender discrepancy between self-report and genotypes (n=7), and 3) aberrant ploidy of sex chromosomes (n=3, one XYY male and two XX/XO mosaic females). B Allele Frequency was used to detect other chromosomal abnormalities.

Quality assessment was performed on 1,008,829 SNPs. SNPs were excluded in two stages. In Stage 1, SNPs were dropped that had less than 95% call rate, and then samples with less than 97% call rates were dropped. Stage 2, SNPs were dropped that had 1) less than 98% call rate 2) any Mendelian errors using HapMap trios (n=583), 3) discordant genotypes using HapMap controls (n=880), 4) discordant genotypes from 2 or more pairs among the 14 study duplicates (n=1765) allowing for 1 error, 5) monomorphic and 6) low minor allele frequency

(MAF < 0.01). SNPs with deviation from Hardy-Weinberg equilibrium (p-value <10E-4) were flagged for future reference but kept in the analysis. The final dataset contained 758,443 SNPs (3,512 with HWE p-value< 10E-4). Of the 2,438 genotyped samples, 2,310 were retained for the association analysis. TSS genotype and phenotype data has been deposited in dbGAP.

VWF Antigen determination:

To create a VWF specific assay, polyclonal rabbit anti-human VWF antibody (DAKO, Glostrup, Denmark) was biotinylated using an NHS activated biotinylating reagent (Solu-link, San Diego, CA). Additionally, non-biotinylated DAKO antibody was conjugated to Alphalisa acceptor beads with sodium cyanoborohydride (Sigma, St Louis, MO). Plasma from the GABC and TSS cohorts were thawed in 37 degree water baths for 15 minutes and mixed by pipette prior to diluting 1:200 in phosphate buffered saline pH 7.4. Five µl of each sample was plated in quadruplicate wells containing 4 µl assay buffer (25 mM HEPES, pH 7.4, 1 mg/mL Dextran 500, 0.5% BSA, 0.5% Triton X-100 and 0.05% Proclin-300), 10 ug/ml VWF antibody-conjugated Alphalisa acceptor beads and biotinylated rabbit anti-human VWF antibody (0.5 nM) using a BioMek FX (Beckman-Coulter, Brea, CA). After incubating for 60 minutes at room temperature, 6 µl (40 µg/ml) of streptavidin coated donor beads (Perkin Elmer) in assay buffer was added and incubated for an additional 30 minutes. The alpha signals were generated and detected on an EnSpire 2300 Multilabel Plate Reader (Perkin Elmer). VWF levels were calculated from the alpha signals of a dilution series of laboratory control plasma with known VWF antigen levels

(FACT). The VWF level for each participant was normalized to the average plate control VWF level. Each sample was independently assayed at least 4 times. The mean sample coefficient of variation was 5% and 3.3% for GABC and TSS, respectively.

Phenotype Data Processing:

The raw VWF levels distribution was normalized by log transformation and adjusted for the effects of BMI, weight, age gender and population structure. The first 10 principal component (PC) scores (see below) were regressed against the age and gender-corrected VWF antigen levels, and the Pearson's correlation coefficients and p-values were extracted. Furthermore, genome-wide association tests were conducted for all SNPs using an additive model and incorporating each PC score separately as a covariate to evaluate its impact on the genomic control factor (GC). After analyzing the Pearson's correlation coefficients, p-values and GC factors, we selected the principal component(s) that were highly correlated with the VWF levels and also resulted in the lowest GC factors. Therefore, log transformed VWF antigen levels were adjusted for age, gender and the selected principal component(s) (PC3 for GABC and no PC adjustment for TSS).

Genetic Analyses

Population Substructure

Due to the presence of sibships of varying size, GABC samples were analyzed using a two-step approach(4). First, SNP eigenvectors and sample eigenscores

were calculated for a subset of 502 unrelated individuals, containing one sample randomly chosen from each sibship. Genotype data for 210 HapMap founder samples representing European (CEU), African (YRI) and Asian (JPT, CHB) populations were included as reference populations. After determining the ancestral information of this population, we ran PCA again without the HapMap samples to avoid the overwhelming signals brought on by them and to generate eigenvectors unaffected by the sibship structure. In the second step, the SNP eigenvectors from the latter PCA run in Step 1 were used to "project" the remaining samples and yield their eigenscores.(4) The first step used a pruned set of ~133k SNPs, obtained from the initial set of 763k autosomal SNPs using an r^2 threshold of 0.2 in windows of 50 SNPs and moving step of 5. Comparisons with the HapMap reference populations revealed 405 unrelated individuals as having European ancestry. These 405 individuals represent 405 families that form a collective set of 940 European subjects that included all the sib-pairs within these families. The initial selection of the random subset of 502 unrelated individuals did not affect the results, as the identical set of European families was obtained when different sets of unrelated subjects were utilized in the first step.

As the majority of the TSS subjects were unrelated, population structure was investigated using a single-step PCA involving all 2,310 subjects and 210 HapMap reference samples and a pruned set of ~117K SNPs, obtained from the initial 758K SNPs using the same pruning criteria described above. All 2,310 subjects were verified as having European ancestry, consistent with their reported Irish ancestry.

Association analyses

We performed single-SNP association tests for the transformed VWF antigen levels using PLINK *(v1.07)*(5), assuming unrelated individuals and an additive mode of allelic effect. To assess the impact of sib relationship, we applied an linear mixed effects model that incorporates the known family structure (R Package, GWAF(6)). In addition, we also applied a GWAS approach based on a mixed linear model taking into account any inferred relatedness and subtle population stratification (EMMAX(7)). For each run, we calculated the genomic control factor(8) to assess the degree of systematic inflation, and p-values were compared between these methods and PLINK in log-scatter plots and QQ plots. The genome-wide significance level was set at $p= 5\times 10^{-8}$ based on Bonferroni Correction for 1 million independent tests.

Meta-analysis

We performed a meta-analysis of GABC and TSS association results using a sample-size-weighted approach to combine the individual association results using METAL(9) on a common set of 723,716 SNPs. A genomic control correction was also applied on the meta-analysis statistics, 1.234 and 1.016 for GABC and TSS, respectively. To examine between-study consistency of effect size and direction we compared regression coefficients for highly associated SNPs in either study.

Linkage Analysis

Linkage analysis was carried out using MERLIN-REGRESS(10). As the signals may be inflated due to unmodeled LD between SNPs (**SI Figure 8**), a clustering

algorithm in MERLIN was adopted to divide SNPs into LD clusters using the threshold of pairwise r^2= 0.001, under the simplifying assumption that there is no recombination within the clusters, and there is no LD between them. This yielded 36,587 clusters; and a linkage LOD score was calculated for each cluster.

To assess the global significance of the cluster-based linkage results, a permutation-based locus-counting approach(11) was implemented. This involved repeating the cluster-based linkage analyses on 1,000 "null" datasets simulated by using the gene-dropping algorithm in MERLIN (conditional on the marker spacing and missingness patterns, family structure and haplotype frequencies) combined with phenotype randomization, and comparing the actual LOD score of each independent region of significant linkage (IRL) with the LOD score null distribution formed by the equal-ranked significant linkage regions across 1000 simulations. Independent regions were defined as the interval extending +/- 20 cM from the SNP/cluster with the local maximum LOD score; thus the distance between the maxima of two adjacent IRLs must be greater than 40 cM. Maximum LOD scores for the top 50 IRLs were recorded in each of the 1,000 runs, and were used to form the 1st-ranked, 2nd-ranked, etc. LOD score null distributions. The actually observed top LOD scores were individually compared with their equal-ranked distribution (**SI Figure 9A**), i.e., the highest LOD score from our study was compared with the 1st-ranked distribution, and the second highest LOD score was compared to the 2nd-ranked distribution, and so on. The original clusters with LOD scores exceeding the 95th percentile of the equal-ranked null distributions were considered significant with empirical $p<0.05$. Starting with the

MERLIN-REGRESS output, the subsequent analysis was carried out using custom scripts in the R programming language.(12)

Variance-Explained by Association and Linkage Regions

The Genome-wide Complex Trait Analysis (GCTA) package(13) was used to estimate the proportion of variance in the VWF levels explained by the entire genome, the top associated SNPs, or intervals representing individual genes or loci. GCTA is a restricted maximum likelihood method (REML) for assessing if individuals with higher genetic similarity also carry a higher phenotypic sharing than is expected by chance. The degree of genetic sharing between all pairs of individuals is computed as a genetic relationship matrix (GRM) using the genotype data for the SNP(s) in question. Then, a linear model is fit to estimate the variance explained by the GRM using the REML approach.

Haplotype-based association analysis

For analyzing haplotype-based association in specific genes (*ABO and VWF*), we phased the genotype data for the entire study (both GABC and TSS) using *BEAGLE* (14), and inferred haplotype blocks in those genes using the phased data and *Haploview* (15). We then performed haplotype association analysis in *PLINK* using the one degree-of-freedom haplotype-specific test for quantitative traits. Haplotypes for the ABO blood type classification (A1, A2, O, B) were tagged by three SNPs: rs8146704, rs8176749, rs687289 as suggested by Barbalic M. et al.(16)

For analyzing haplotype association in the Chr2 linkage region we focused on the GABC study, as it contained most of the familial samples and contributed most to

the linkage signal. In the first approach, we phased the genotype data for all European individuals in GABC (n=940) in the Chr2 linkage region plus ~20 Mb flanking regions (60 Mb to 125 Mb) using *BEAGLE(17)*, and defined haplotype blocks using --blocks command in *PLINK*(5), which follows the default procedure in *Haploview*. We then performed haplotype association analysis in *PLINK* for every block in the linkage region. Since the results included a p-value for every major haplotype (defined by haplotype frequency> 1%) in a given block, we extracted the smallest p-value (i.e., for the haplotype with the strongest association) to represent each block. To compare with single-SNP association results we similarly extracted the minimum p-value among the SNPs residing in each haplotype block.

In the second approach, we analyzed the differences in LD structures between high-VWF and low-VWF individuals around SNPs with moderately strong association with VWF (i.e., SNP with $p<0.001$) in the Chr2 linkage region, following the approach of Wang et al. (18). Unlike the first approach, which tests association with common haplotypes, the scenario we are testing in the second approach is that if the causal variants are rare and that they arose recently, they would likely be on specific long-range haplotypes that exist more frequently in the individuals with high-VWF than in individuals with low-VWF, or vice versa. We defined the high-VWF group as those falling in the upper 25% of the VWF distribution (n=241), and likewise the low-VWF group as those falling in the lowest 25% (n=241), essentially creating a case-control contrast between the two tails of a quantitative distribution. Since we were searching for alleles that either

increase or decrease VWF levels we examined enrichment of long-range haplotypes both in high-VWF and in low-VWF groups. There were 11 SNPs with single-SNP association p-value < 0.001 in the linkage region in Chr2: 74.98 Mb-108.95 Mb, and they form eight candidate association regions, defined as the 1Mb intervals (500 kb on each side) flanking the "index SNPs". The goal of subsequent steps was to find a subset of genotyped SNPs with different levels of LD strength, measured by D', between the high-VWF and low-VWF groups. To do so, we first selected SNPs in each 1Mb interval having p<0.1 in the single-SNP association test in GABC, yielding 3 to 75 SNPs across the eight intervals, spanning 305 Kb to 987 Kb. For each of these SNPs, we calculated D' to the index SNP for the high- and low-VWF groups separately, and calculated the ratio of the D' in the high-VWF group and the D' in the low-VWF group as a measure of differential LD structure. As stated above, the unknown causal variant could be tagged by long-range haplotypes formed by the SNPs with large D' ratios. In all, we ran 16 tests, for eight intervals and two types of D' differences (D' ratio >=2 or <=0.5), and found that six tests yielded at least three SNPs passing the indicated D' ratio (D' ratio >=2 and/or <=0.5) around eight index SNPs with p<0.001 in the Chr2 linkage region. These six SNP sets, corresponded to the following index SNPs and D' ratios: rs9808242 (pos=85,140,799, D' ratio >=2 and <=0.5), rs6547231 (pos=76,719,093, D' ratio >=2 and <=0.5), rs7566719 (pos=76,719,093, D' ratio < 0.5), and rs1543282 (pos=78,118,961, D' ratio < 0.5). Treating each of the six sets of SNPs as a pre-defined long-range haplotype block, we analyzed haplotype-based test of VWF association using phased

genotype data for the 940 individuals. To assess the region-wide significance of these association results, we carried out 1,000 simulations by randomizing the phenotypes among the 940 subjects, and running haplotype association using the same six sets of selected SNPs. Since each SNP set produced multiple association p values, one for each haplotype, we extracted the minimum p-values within each set during each of the 1000 runs, and calculated an empirical p-value for the original minimum p-values in each set by ranking it in the null distribution formed by the 1000 minimum p-values.

Supplemental References

1. Laurie CC*, et al.* (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* 34(6):591-602.

2. GENEVA Study members. (2012) Geneva study data cleaning reports. Available @ https://www.genevastudy.org/Data_Cleaning_Reports. Last accessed 11-14-2012.

3. Desch K, Li J, & Ginsburg D (2011) dbGaP: The Genes and Blood Clotting Study. Available @ http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000304.v1.p1. Last accessed 11-14-2012.

4. Zhu X, Li S, Cooper RS, & Elston RC (2008) A unified association analysis approach for family and unrelated samples correcting for stratification. *American journal of human genetics* 82(2):352-365.

5. Purcell S*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 81(3):559-575.

6. Chen MH & Yang Q (2010) GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 26(4):580-581.

7. Kang HM*, et al.* (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348-354.

8. Devlin B & Roeder K (1999) Genomic control for association studies. *Biometrics* 55(4):997-1004.

9. Willer CJ, Li Y, & Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26(17):2190-2191.

10. Abecasis GR, Cherny SS, Cookson WO, & Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30(1):97-101.

11. Wiltshire S, Cardon LR, & McCarthy MI (2002) Evaluating the results of genomewide linkage scans of complex traits by locus counting. *American journal of human genetics* 71(5):1175-1182.

12. Ihaka R & Gentleman R (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3):299-314.

13. Yang J, Lee SH, Goddard ME, & Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* 88(1):76-82.

14. Browning BL & Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84(2):210-223.

15. Barrett JC, Fry B, Maller J, & Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2):263-265.

16. Barbalic M*, et al.* (2010) Large-scale genomic studies reveal central role of ABO in sP-selectin and sICAM-1 levels. *Hum Mol Genet* 19(9):1863-1872.

17. Browning BL & Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American journal of human genetics* 84(2):210-223.

18. Wang K*, et al.* (2010) Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet* 86(5):730-742.

Supplementary Table 1. Top GABC SNPs. Genome-wide significant (p-value < 5.0E-8) SNPs in GABC association analysis, sorted by p-value, with respect to their relationship to the nearest gene.

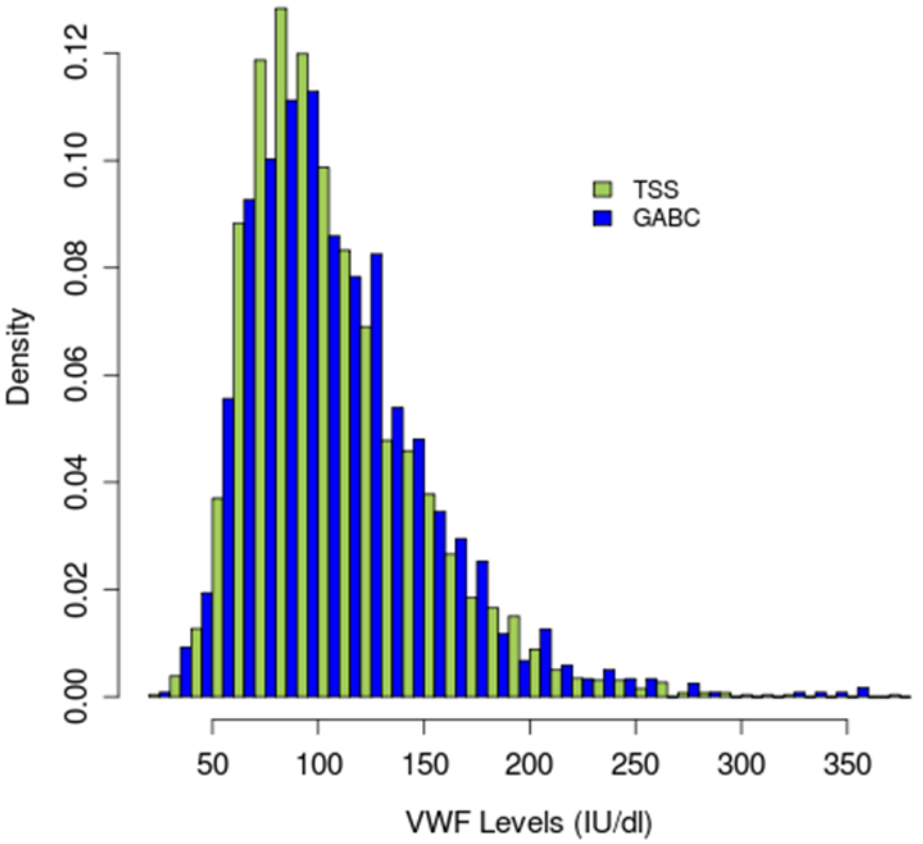
REGION	N SNPs	TOP SNP	POSITION	CODED ALLELE	CLOSEST GENE	ALLELE FREQUENCY	BETA (SE)	P-VALUE
9q34	31	rs687289	135126927	A	*ABO*	0.357	0.36 (0.022)	1.7E-52
9q34	1	rs11244035	135071140	T	*OBP2B*	0.115	0.27 (0.039)	3.3E-12
9q34	2	rs3094379	135324731	T	*CACFD1*	0.161	0.21 (0.032)	1.2E-10
9q34	1	rs45618736	137580983	A	*OBP2A*	0.110	0.26 (0.040)	1.3E-10
9q34	1	rs4962153	135313575	A	*ADAMTS13*	0.161	0.21 (0.032)	1.6E-10
9q34	1	rs11244079	135174347	A	*LOC653163*	0.051	0.33 (0.055)	4.6E-09
9q34	1	rs7868232	135032085	C	*GBGT1*	0.256	0.15 (0.027)	2.8E-08

Supplementary Table 2. Top TSS SNPs. Genome-wide significant (p-value < 5.0E-8) SNPs in TSS association analysis, sorted by p-value, with respect to their relationship to the nearest gene.

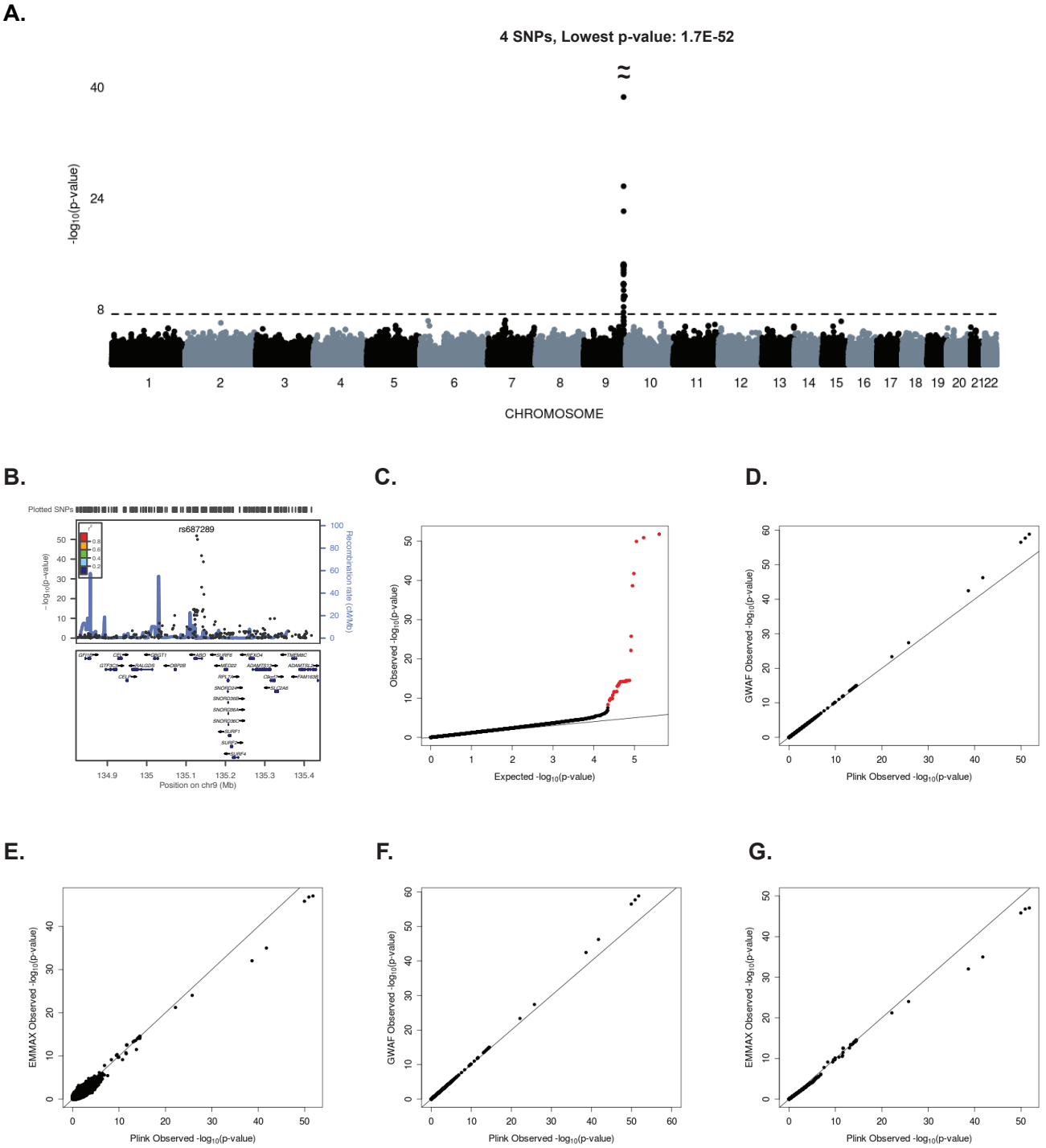
REGION	N SNPs	TOP SNP	POSITION	CODED ALLELE	CLOSEST GENE	ALLELE FREQUENCY	BETA (SE)	P-VALUE
9q34	31	rs687289	135126927	T	*ABO*	0.249	0.33 (0.016)	3.7E-89
9q34	1	rs11244079	135174347	A	*LOC653163*	0.0630	0.35 (0.030)	3.9E-31
9q34	2	rs4962153	135313575	A	*ADAMTS13*	0.106	0.25 (0.024)	1.2E-24
9q34	11	rs7855713	135102562	G	*LOC286310*	0.109	0.22(0.024)	7.4E-20
9q34	7	rs11244035	135071140	T	*OBP2B*	0.0851	0.23 (0.027)	3.4E-17
9q34	1	rs45618736	137580983	A	*OBP2A*	0.0822	0.24 (0.029)	5.4E-17
9q34	1	rs28602591	135236487	T	*C9orf96*	0.0716	0.24 (0.029)	4.8E-16
12p13	10	rs1063856	6023795	G	*VWF*	0.395	0.12 (0.015)	8.5E-14

Supplemental Table 3. The top two haplotype-based association signals. Shown are the raw p values, frequencies, and permutation-based empirical p values for the strongest associated haplotypes in two of the SNP sets, anchored by rs7566719 and rs6547231, respectively.

Significant Interval	P-Value	Freq	Empirical P-Value of the Interval
Interval 1 (rs7566719)	1.9E-04	0.02	2.0E-03
Interval 2 (rs6547231)	4.4E-04	0.28	6.0E-03

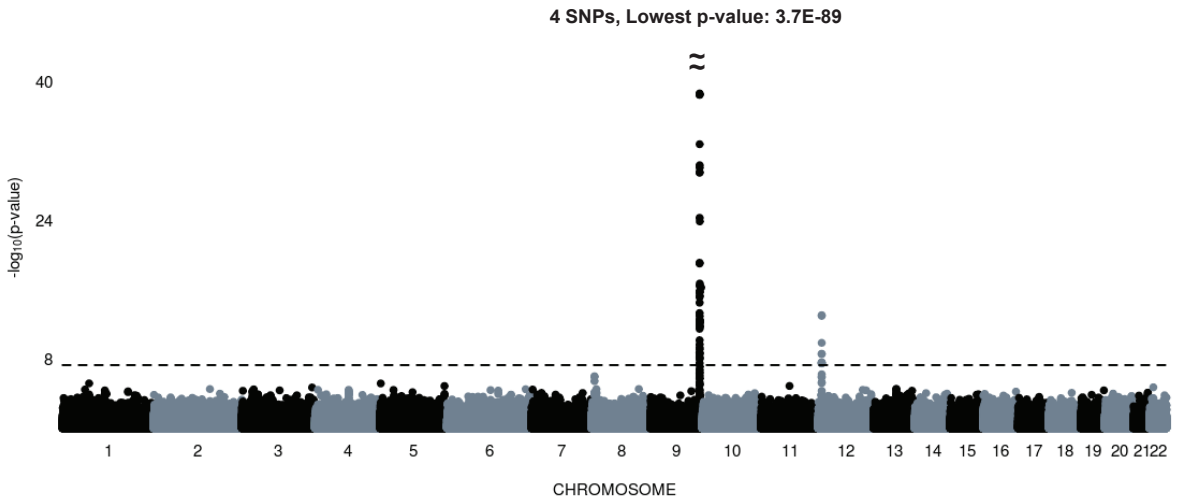


Supplemental Figure 1. Untransformed VWF level distributions of the GABC and TSS datasets. Percent of individuals in VWF level bins are interleaved by cohort.

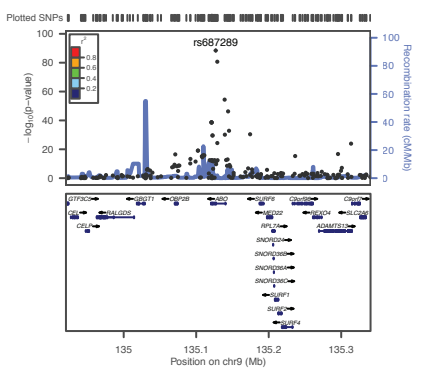


Supplementary Figure 2. GABC association analysis (~763K SNPs) using age-, gender-, and principal component-adjusted VWF values. **A**. Genome-wide -log10 (p-values) plot. The horizontal line marks the 5.0 E-8 threshold of genome-wide significance. Peak at Chromosome 9 is within the ABO gene. Four additional SNPs with p-values < 1E-40 in this peak are not plotted. **B**. Regional plot for the ABO gene on chromosome 9. **C**. Quantile-Quantile plot of observed vs. expected –log10(p-values) for VWF association without any correction for genomic control (Observed p-values < 5.0E-8 are in red). **D**. Scatter plot comparison of the observed –log10(p-values) using Plink and GWAF. **E**. Scatter plot comparison of the observed –log10(p-values) using Plink and EMMAX. **F**. Quantile-Quantile plot comparison of the observed –log10(p-values) using Plink and GWAF. **G**. Quantile-Quantile plot comparison of the observed –log10(p-values) using Plink and EMMAX.

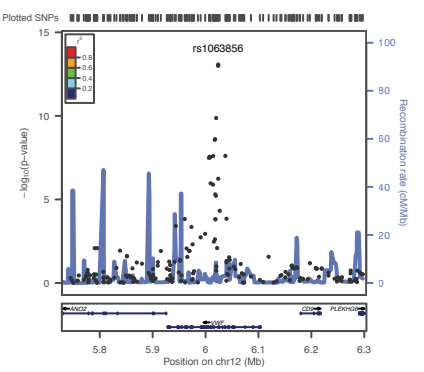
A.



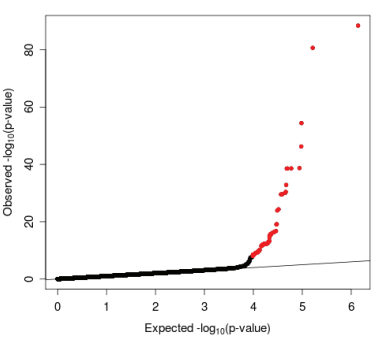
B.



C.

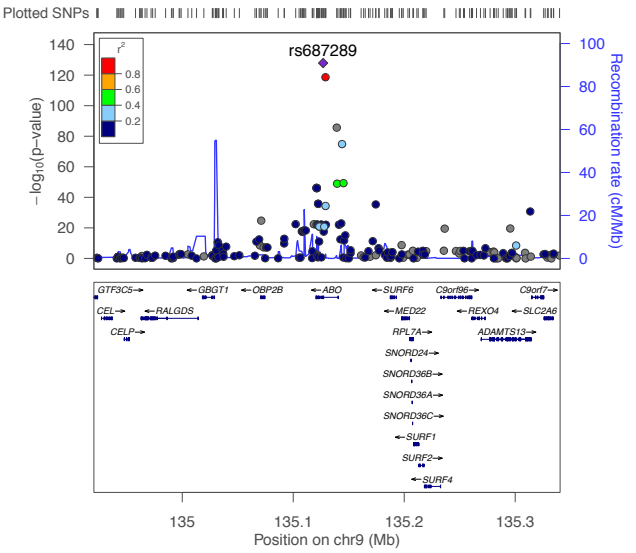


D.

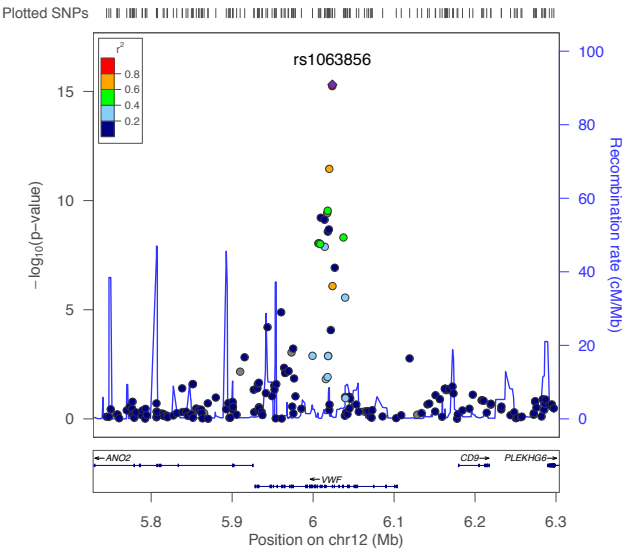


Supplementary Figure 3. TSS association analysis (~739K SNPs) using age-, and gender-adjusted VWF values. **A**. Genome-wide -log10 (p-values) plot. The horizontal line marks the 5.0E-8 threshold of genome-wide significance. Peaks at Chromosome 9 and Chromosome 12 are within the ABO and VWF genes, respectively. Four additional SNPs with p-values < 1E-40 in the Chromosome 9 peak are not plotted. **B**. Regional plot for the ABO gene on Chromosome 9. **C.** Regional plot for the VWF gene on chromosome 12. **D.** Quantile-Quantile plot of observed vs. expected –log10(p-values) for VWF association without any correction for genomic control (Observed p-value< 5.0E-8 are in red).

A.

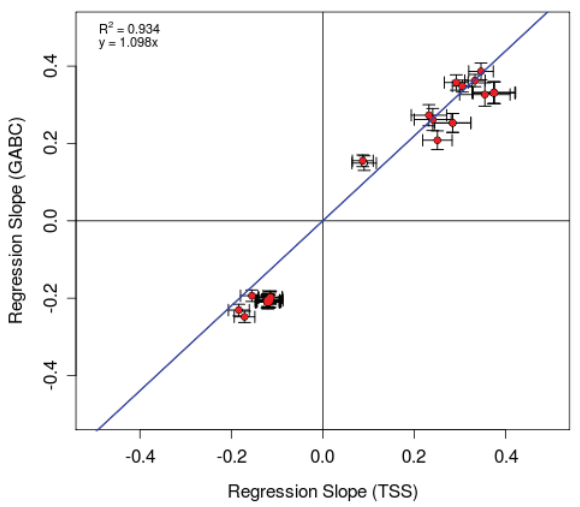


B.

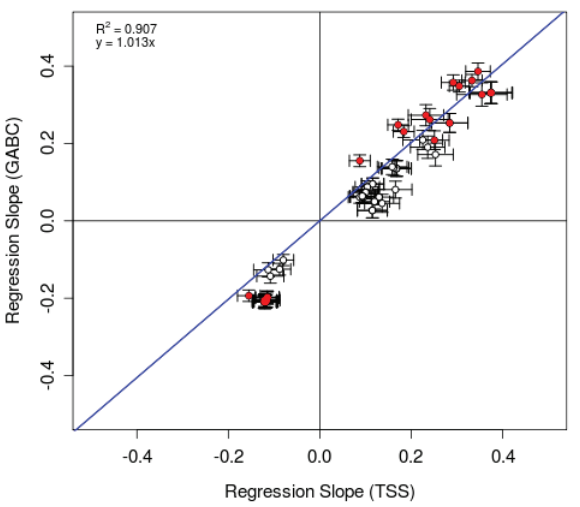


Suplemental Figure 4. Meta-analysis results. **A.** Regional plot for the ABO gene on Chr9. **B.** Regional plot for the VWF gene on Chr12.

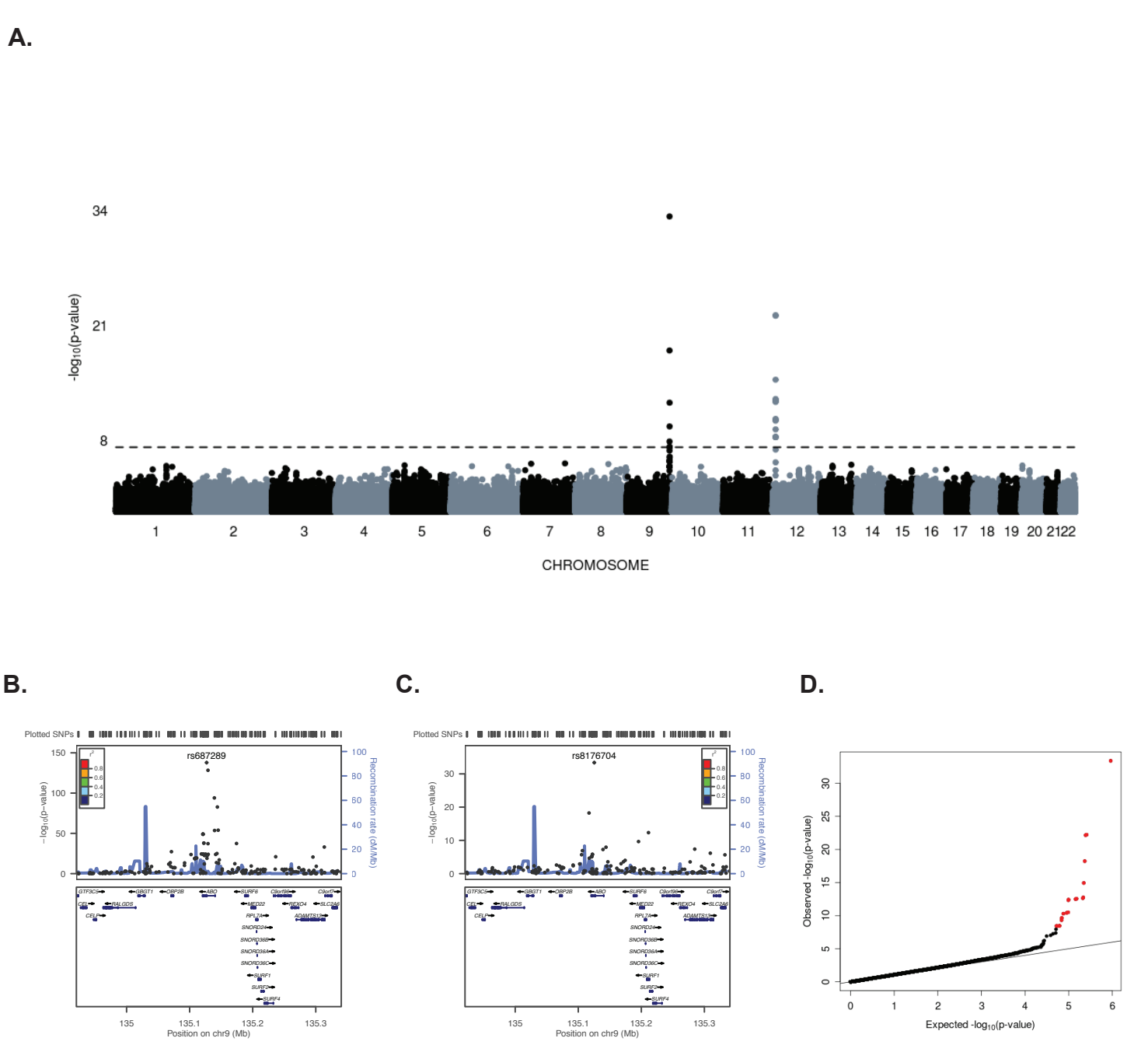
A.



B.

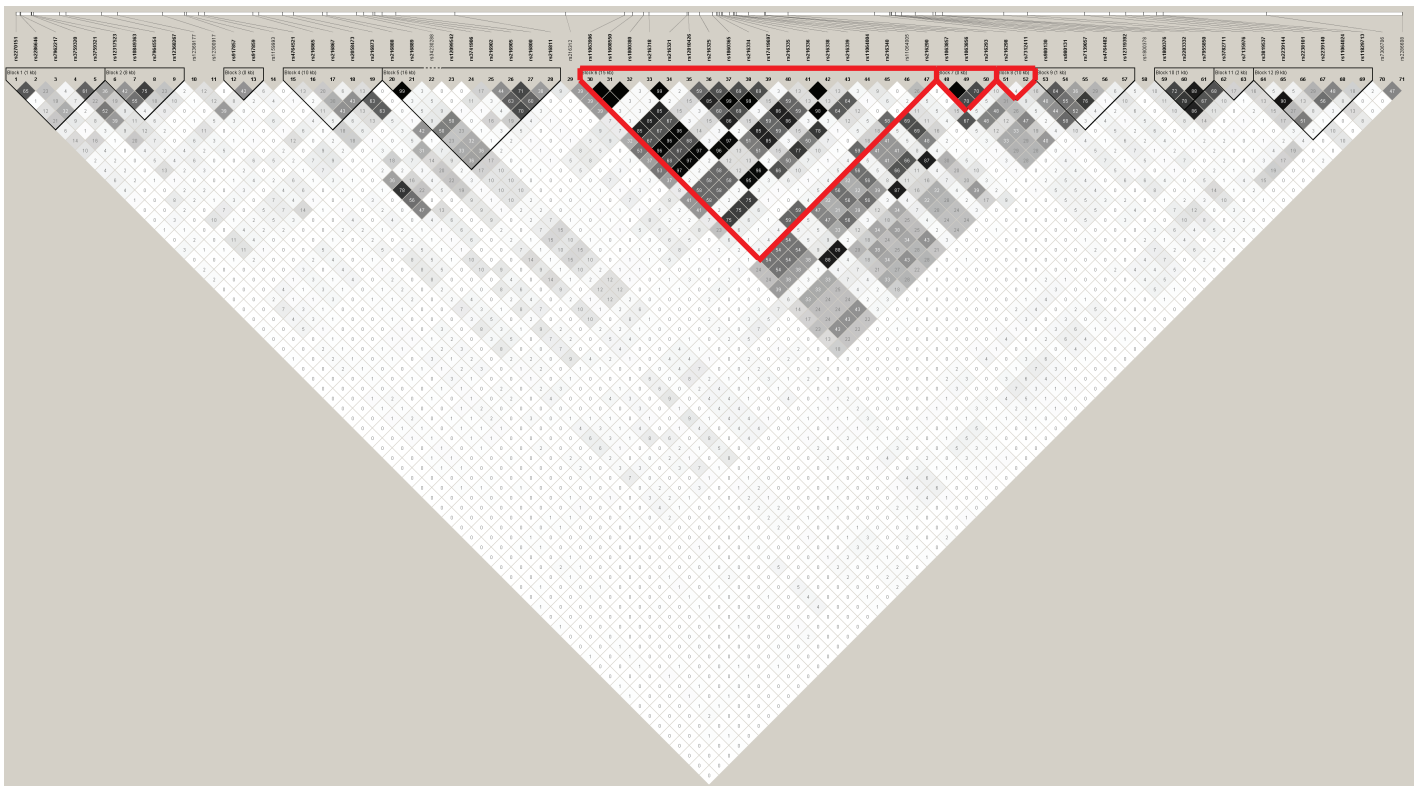


Supplementary Figure 5. Comparisons of effect size and direction between GABC and TSS for **A.** top 38 SNPs in GABC and **B.** top 64 SNPs in TSS. Shown are beta values (i.e., regression slopes) for TSS (x axis) and GABC (y axis). Error bars represent the standard errors of the estimated beta values in each dataset. Marked in red are SNPs that are significant in both GABC and TSS, for 34/38 and 34/64 SNPs in A and B, respectively.

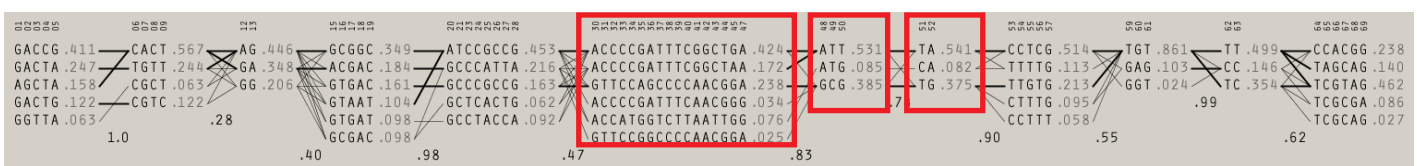


Supplementary Figure 6. GABC+TSS conditional association analysis (~724k SNPs) using age-, gender-, and principal component-adjusted VWF values after previously identified Top SNP (rs687289) is used as a covariate. **A**. Genome-wide -log10 (p-value) plot. The horizontal line marks the 5 ×10-8 threshold of genome-wide significance. Peaks at Chromosome 9 and Chromosome 12 are within the ABO and VWF genes, respectively. **B.** Regional plot for the ABO gene on Chromosome 9 (before rs687289 is used as a covariate). **C.** Regional plot for the ABO gene on chromosome 9 (after rs687289 is used as a covariate). **D.** Quantile-Quantile plot of observed vs. expected –log10(p-values) for VWF conditional association without any correction for genomic control (Observed p-values< 5.0E-8 are in red).

A.

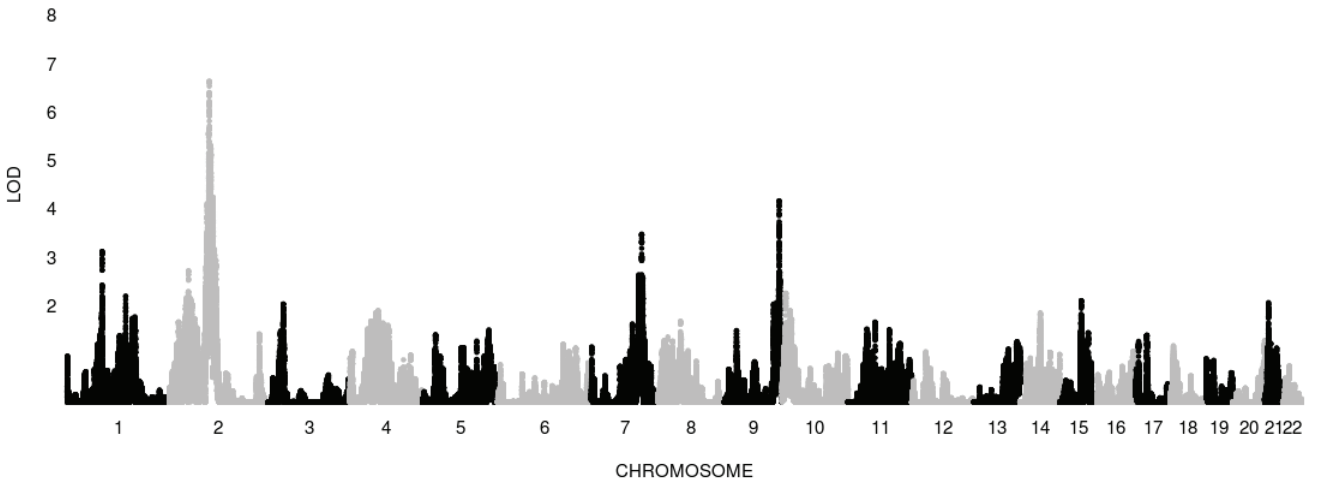


B.

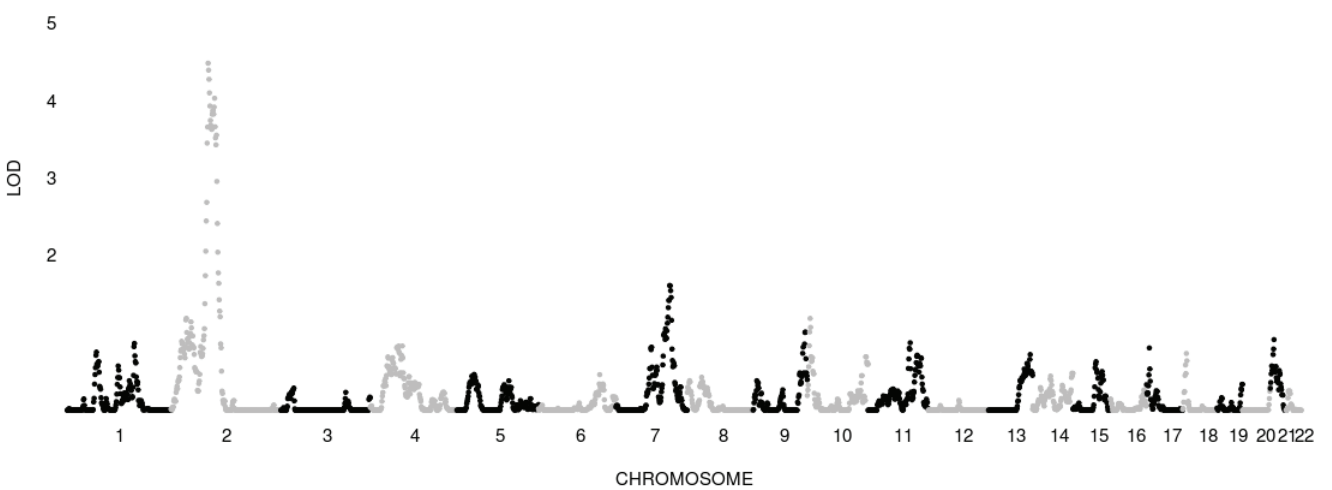


Supplementary Figure 7. A. Haplotypes in the VWF gene constructed by Haploview program using the default settings and a combined set of GABC (n=940) and TSS (n=2310). 7 LD blocks are identified. 15 of the 73 significant SNPs in the meta-analysis are distributed within the Blocks 1, 2 and 3, which are marked in red. **B**. Frequencies of the most common haplotypes in each of the LD blocks. Blocks 1, 2 and 3 are marked in red.

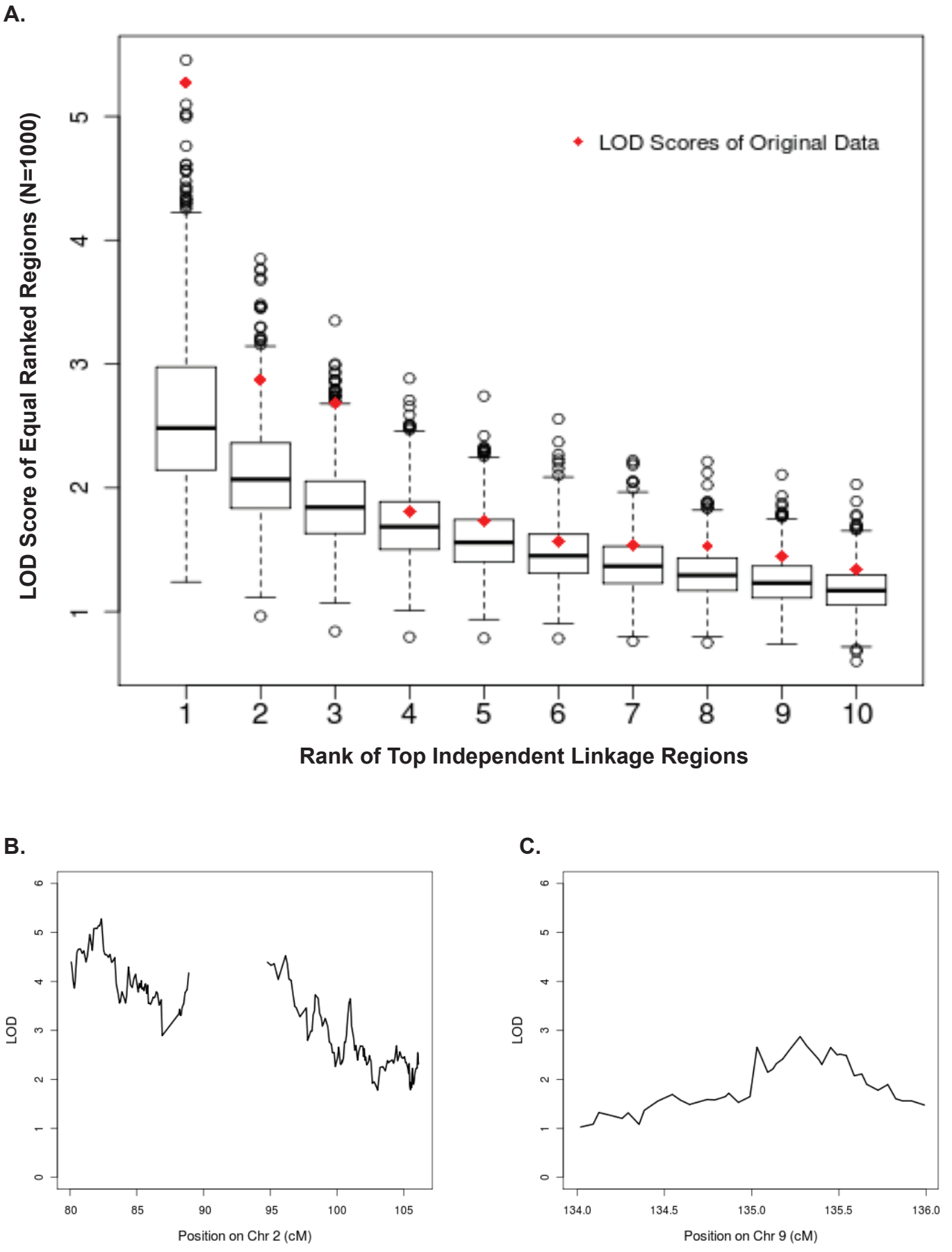
A.



B.

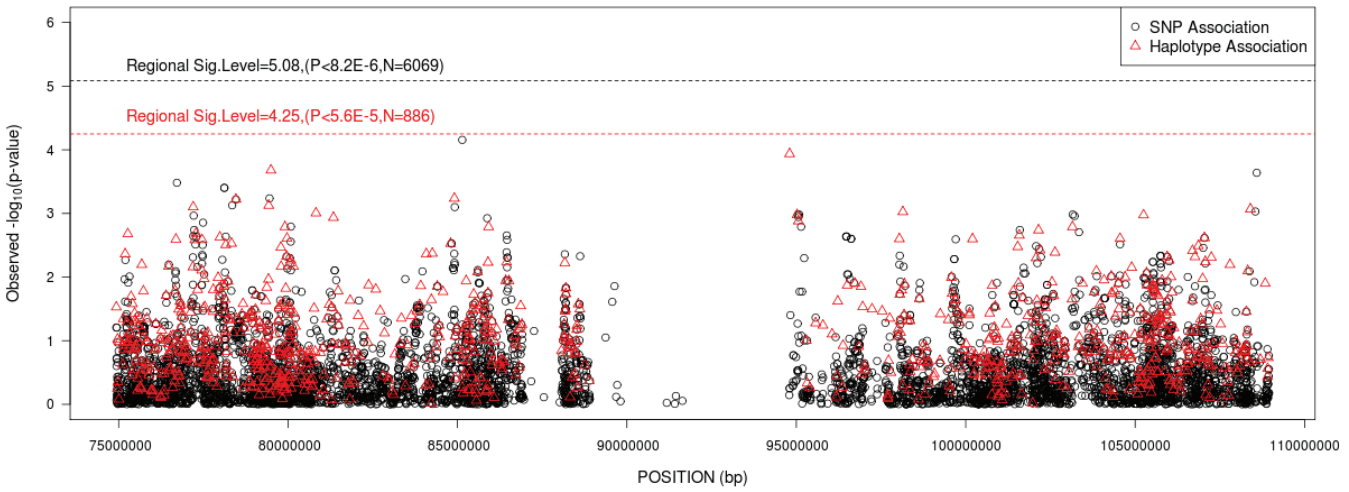


Supplementary Figure 8. A. LOD scores for the GABC + TSS linkage analysis using all SNPs (~760k). **B.** LOD scores for the GABC + TSS linkage analysis using a pruned set of 2.7K SNPs (criterion: 1 SNP every 1Mb, picking the SNP with the largest MAF in each window)

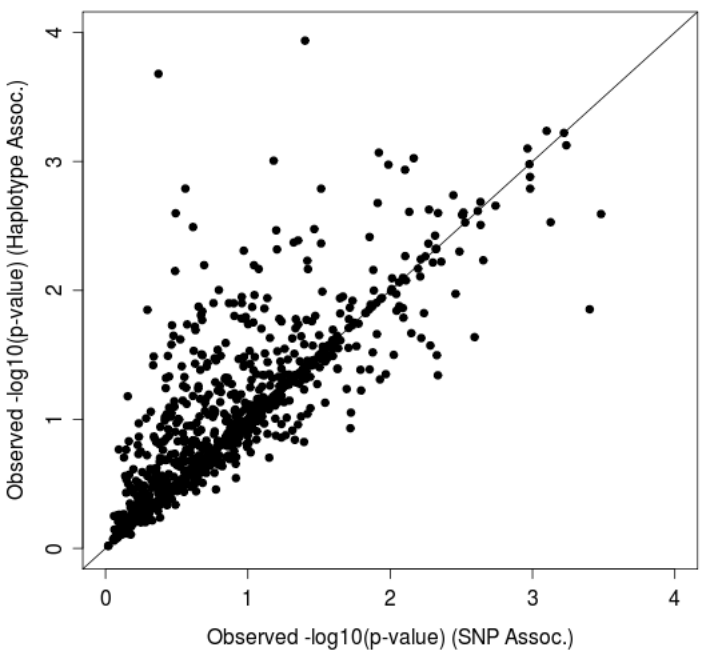


Supplemental Figure 9. **A.** Comparison of the observed maximal LOD scores in the top 10 Independent Linkage Regions (shown as red dots) with their corresponding equal-ranked LOD score distributions in 1,000 null simulations (shown as boxplots). Independent Linkage Regions have LOD score maxima separated by >40 cM from each other. Null simulations are created by randomizing the phenotypes among all individuals as well as modeling LD using the previously defined clusters. **B.** Regional LOD score plot for the linkage peak on Chr2 (74.93Mb-108.95Mb). **C.** Regional LOD score plot for the linkage peak on Chr9 (135.03Mb-135.82Mb).

A.



B.



Supplemental Figure 10. Haplotype Association in Chr2 linkage interval. A. –log(p-value) plot of single-SNP association (black) and haplotype-based association (minimal haplotype p-value, red) in the Chr2 linkage region. **B.** Comparison of the minimal haplotype-based p values (y axis) and the minimal per-SNP p-values across 866 haplotype blocks in the Chr2 linkage region.