Table S1: Numbers of samples and SNPs in the final analysis.

	CAD	CD	RA	T1D	T2D
Cases	1,856	1,566	1,739	1,912	1,806
Controls	2,826	2,770	2,737	2,869	2,729
SNPs	394,589	395,452	395,003	394,681	394,951

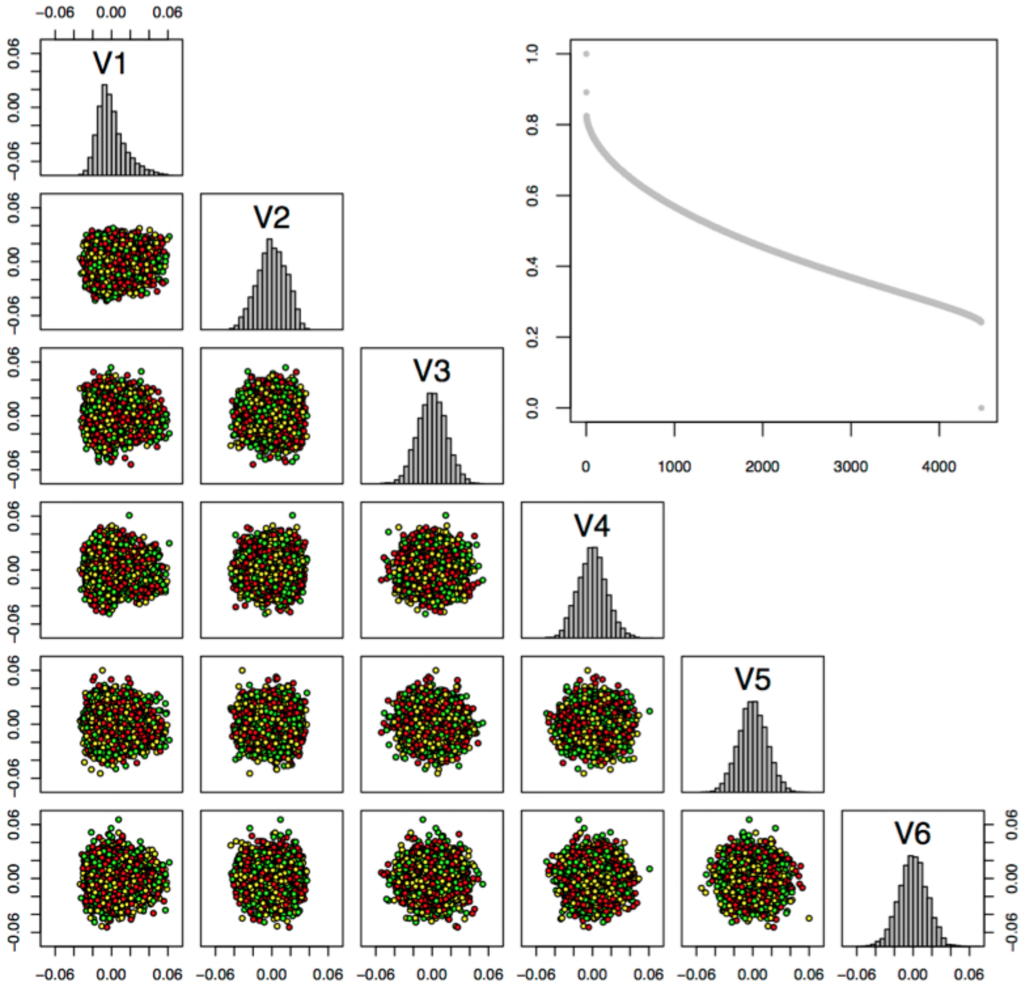

Figure S1: Pairwise plots of top 6 eigenvectors, their histogram, and all eigenvalues (scaled according to the largest eigenvalue) of rheumatoid arthritis data. Red dots denote cases, green dots denote 58BC controls and yellow dots denote NBS controls.

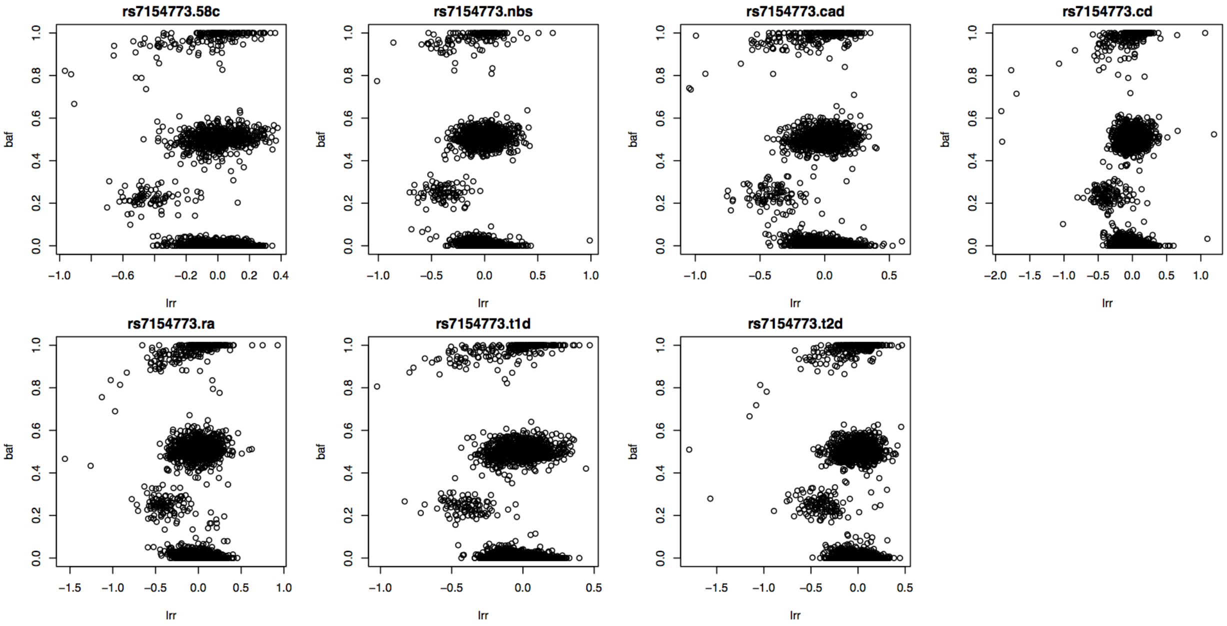

Figure S2: Cluster plot for SNP rs7154773 for two controls and five disease cases. There is a fourth cluster in all seven samples. In each panel, the x-axis is the logR-Ratio (LRR), and the y-axis is the B-allele frequency (BAF). The LRR is a normalized measure of the total signal intensity for two alleles of the SNP. The BAF is a normalized measure of the allelic intensity ratio of two alleles.

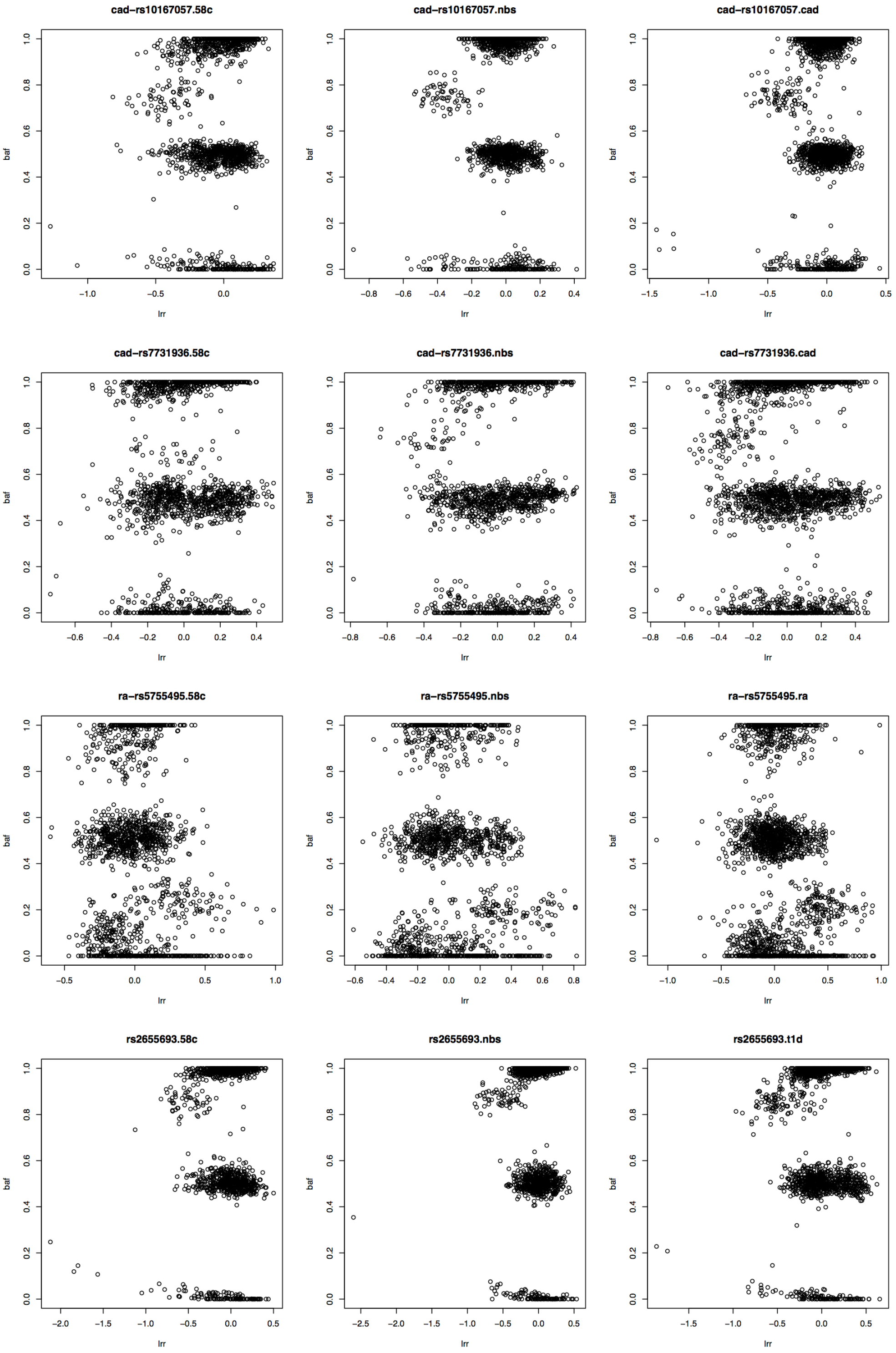

Figure S3: Cluster plots for core SNPs of orphan signals.

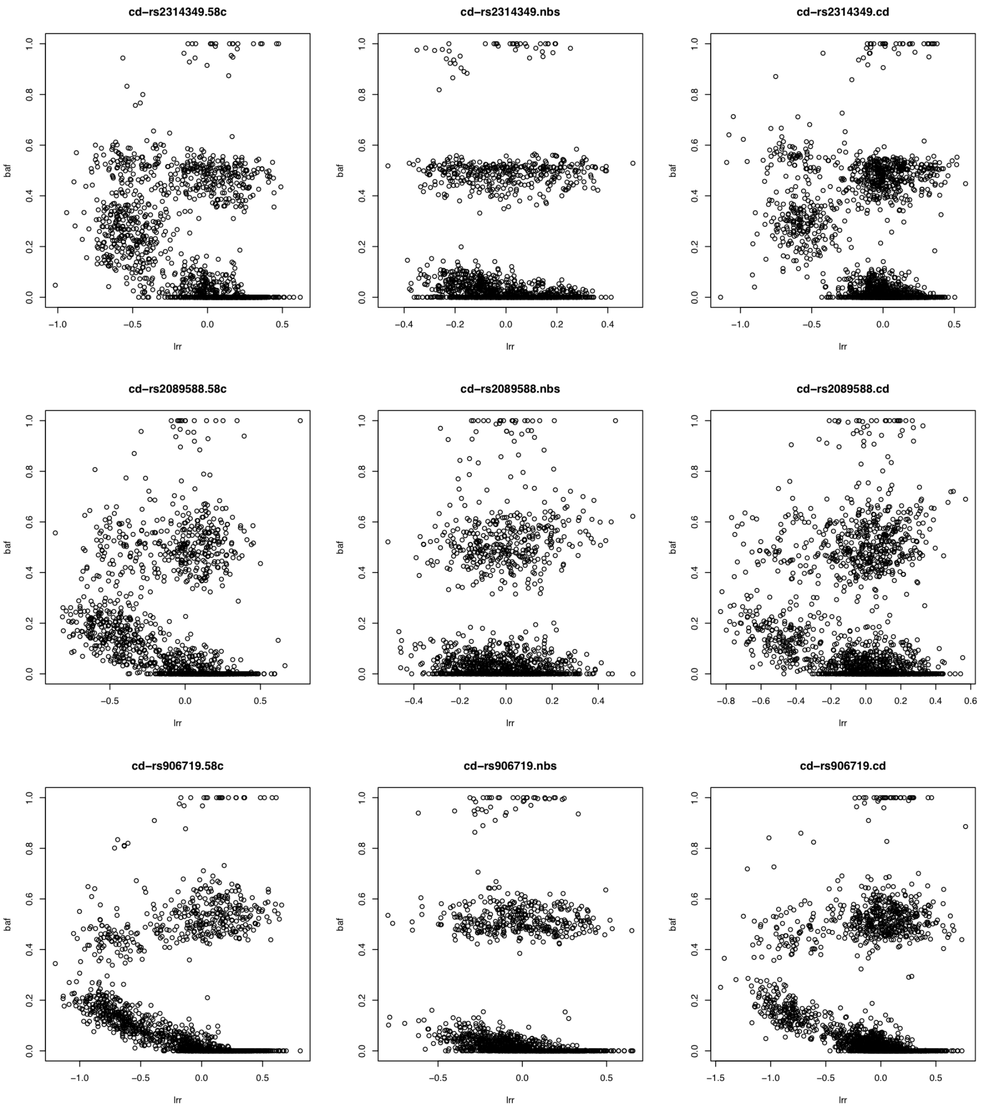

Figure S4: Cluster plots for three SNPs in gene *MCF2L2* (on chromosome 3), which shows a strong association with Crohn's disease. After these three SNPs were removed, the signal disappeared.

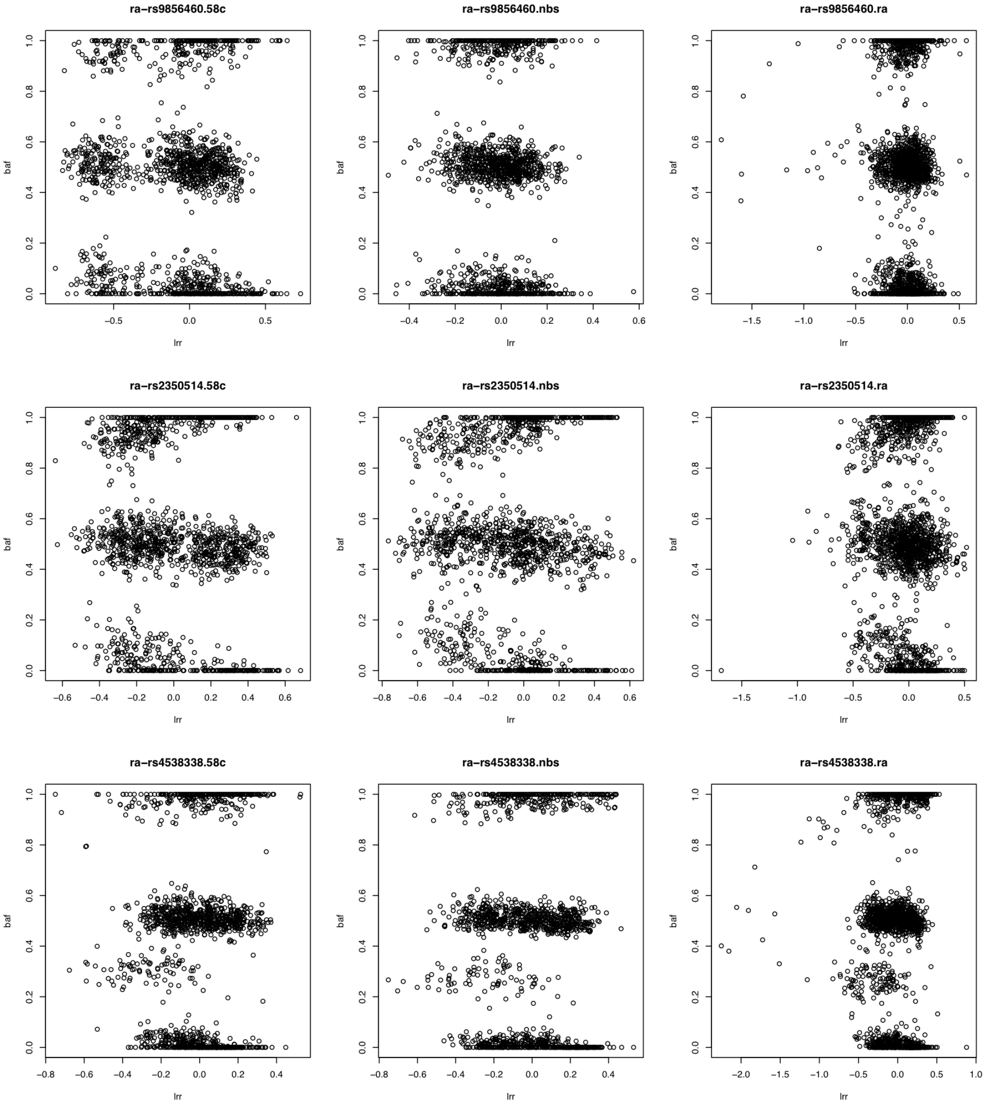

Figure S5: Cluster plots for three SNPs in gene *CLSNT2* (on chromosome 3), which shows a strong association with rheumatoid arthritis. After these three SNPs were removed, the signal disappeared.

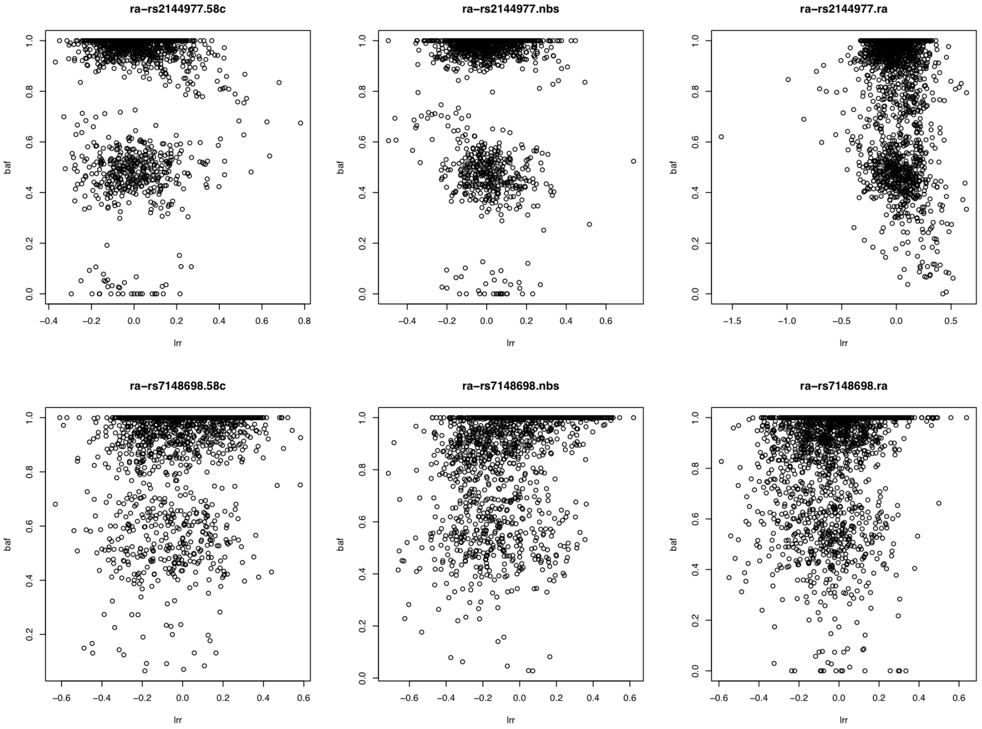

Figure S6: Cluster plots of two SNPs near gene *NID2* (on chromosome 14), which shows a strong association with rheumatoid arthritis. After these two SNPs were removed, the signal disappeared.