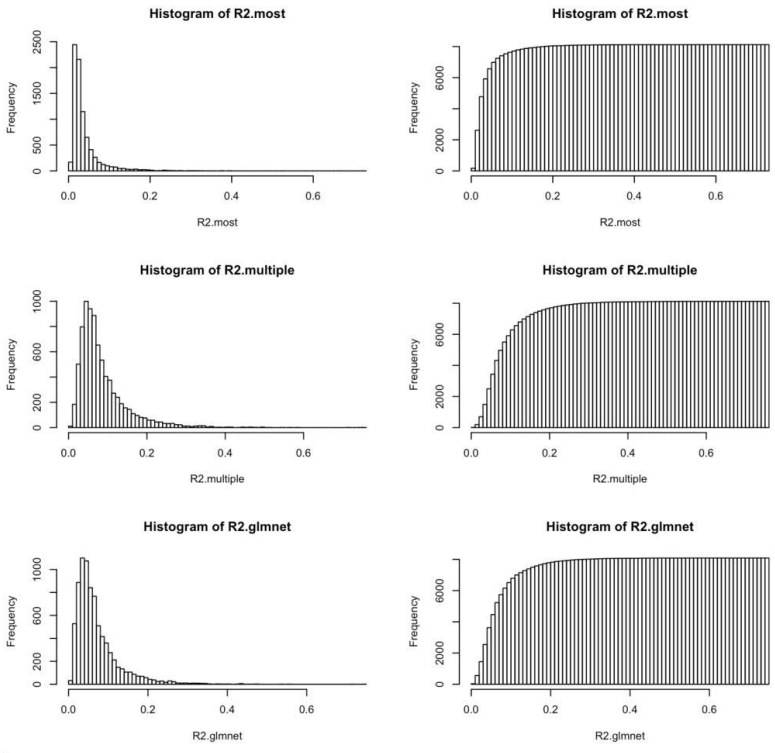
Supplementary Table 1. Numbers of probes and genes that have LASSO models

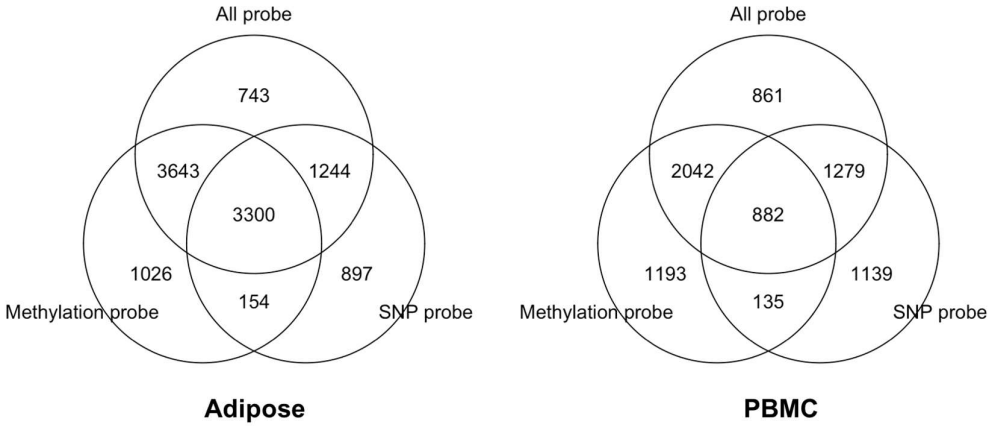
Datasets		Probes		Genes	
		w/LASSO model	w/o LASSO model	w/LASSO models	w/o LASSO model
Adipose	Methylation Probes	150349	275668	4252	24972
	SNP/CrossHybr probes	46119	115291	5595	17078
	Total	213764	371635	8930	17806
PBMC	Methylation Probes	73553	404434	4252	24972
	SNP/CrossHybr probes	28345	152568	3435	22808
	Total	107484	541550	5064	25966

Tables shows the number of CpG probes and the number of genes used in the LASSO regression models. Three kinds of LASSO regression models were performed using different categories of CpG probes (using only DNA methylation probes, using probes with SNP or cross-hybridization probes, and using all the probes). 'w/LASSO model' indicates the number of probes or genes with LASSO model, 'w/o LASSO model' shows the number of probes or genes failed to perform LASSO model.

Supplementary Figure 1. Histograms of R^2 from fitting different regression models. R2.most is R2 from single regression, R2.multiple is R2 from multiple regressions, and R2.glmnet is R2 from lasso regression.



Supplementary Figure 2. Number of genes with LASSO model from different types of probes. The left panel is the Venn diagram for Adipose dataset, and the right is for PBMC dataset. All, all the probes; Methylation, methylation probes; S&C, SNP and cross hybridization probes.



Supplementary Table 2. Gene Ontology enrichment analysis for genes with predictive R^2 greater than 0.1 in PBMC dataset.

Term	Fold Enrich	Pvalue	FDR
GO:0006952~defense response	2.61	2.41E-07	4.14E-04
GO:0002684~positive regulation of immune system process	4.13	3.96E-08	6.81E-05
GO:0046649~lymphocyte activation	4.28	2.17E-07	3.73E-04
GO:0001775~cell activation	3.43	9.96E-07	1.71E-03
GO:0045321~leukocyte activation	3.69	1.08E-06	1.85E-03
GO:0042110~T cell activation	5.04	1.55E-06	2.66E-03
GO:0002694~regulation of leukocyte activation	4.34	1.89E-06	3.25E-03
GO:0051249~regulation of lymphocyte activation	4.58	2.12E-06	3.64E-03
GO:0050865~regulation of cell activation	4.11	3.78E-06	6.50E-03

The p-values are the probability of observing the number of genes out of total genes in a particular term given the number of genes in the whole genome belonging to this term.

Supplementary Table 3. Gene Ontology and pathway enrichment analysis for genes with predictive R^2 greater than 0.1 in Adipose dataset using all the probes.

Term	Fold Enrich	P-value	FDR
GO:0006952~defense response	2.31	4.31E-11	7.79E-08
GO:0009611~response to wounding	2.21	1.12E-08	2.03E-05
GO:0006954~inflammatory response	2.53	7.22E-08	1.30E-04
GO:0002504~antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	14.99	1.17E-13	2.11E-10
GO:0030036~actin cytoskeleton organization	2.43	3.20E-05	5.78E-02
KEGG_PATHWAY:hsa05330:Allograft rejection	6.67	2.48E-09	3.02E-06
KEGG_PATHWAY:hsa05332:Graft-versus-host disease	6.16	9.30E-09	1.14E-05
KEGG_PATHWAY:hsa04612:Antigen processing and presentation	4.16	1.04E-08	1.27E-05

KEGG_PATHWAY:hsa04940:Type I diabetes mellitus	5.72	3.03E-08	3.71E-05
KEGG_PATHWAY:hsa05416:Viral myocarditis	4.35	5.54E-08	6.76E-05

The p-values are the probability of observing the number of genes out of total genes in a particular term given the number of genes in the whole genome belonging to this term.

Supplementary Table 4. Gene Ontology and pathway enrichment analysis for genes with predictive R^2 greater than 0.1 in PBMC dataset using all the probes.

Term	Fold Enrich	P-value	FDR
GO:0006952~defense response	2.62	7.90E-11	1.38E-07
GO:0006954~inflammatory response	2.55	1.02E-05	1.77E-02
GO:0002684~positive regulation of immune system process	3.10	1.02E-06	1.78E-03
GO:0051249~regulation of lymphocyte activation	3.42	1.92E-05	3.36E-02
GO:0002684~positive regulation of immune system process	3.10	1.02E-06	1.78E-03
GO:0046649~lymphocyte activation	2.99	2.47E-05	4.32E-02
KEGG_PATHWAY:hsa05332:Graft-versus-host disease	8.16	1.03E-07	1.21E-04
KEGG_PATHWAY:hsa05330:Allograft rejection	8.11	4.70E-07	5.53E-04
KEGG_PATHWAY:hsa05416:Viral myocarditis	5.00	7.36E-06	8.65E-03
KEGG_PATHWAY:hsa05320:Autoimmune thyroid disease	5.72	1.45E-05	1.70E-02

The p-values are the probability of observing the number of genes out of total genes in a particular term given the number of genes in the whole genome belonging to this term.