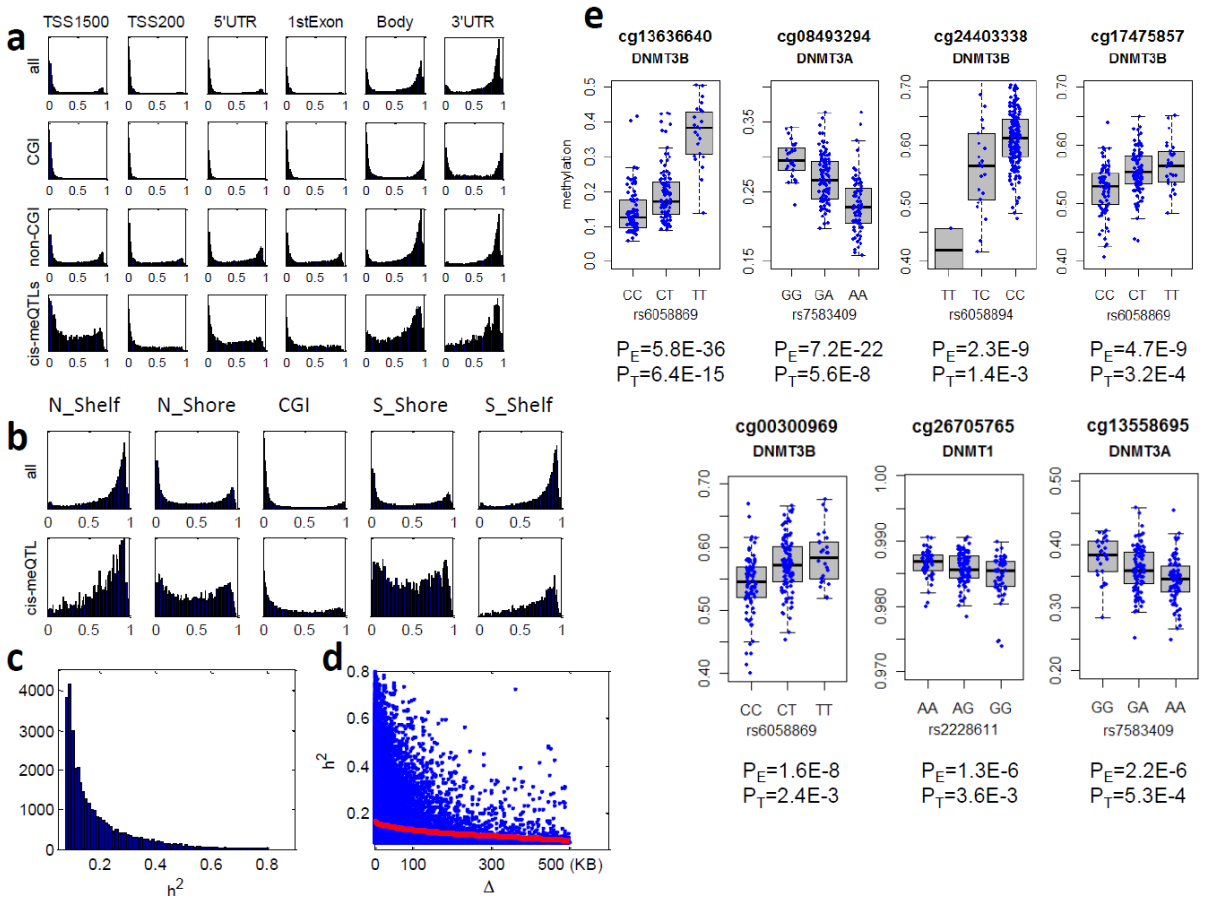
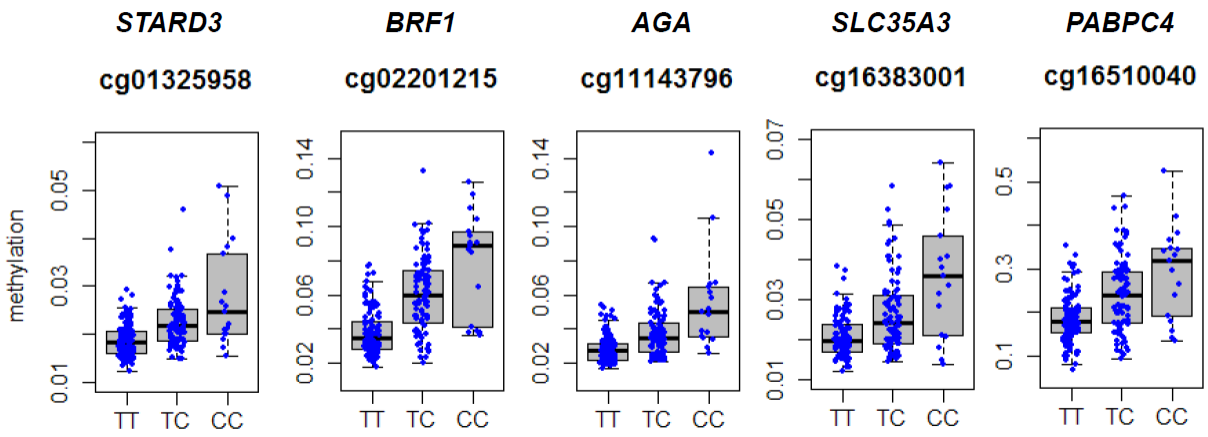
Supplementary Information

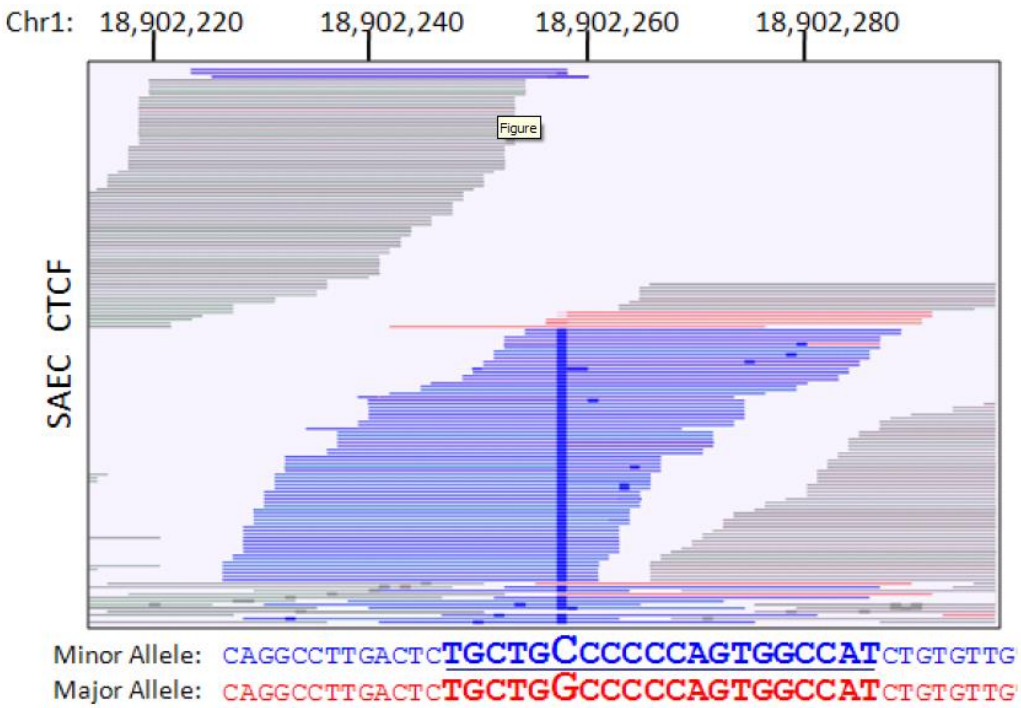
Index of Supplementary Materials



Supplementary Figure 1. Distribution of methylation levels, characteristics of *cis*-meQTLs and association plots for meQTLs in methylation-related genes. (a) For each methylation probe, we computed the average methylation level across 210 samples. The histograms show the distributions of average methylation levels in different categories. CGIs in promoter regions are primarily unmethylated. Non-CGI regions are more likely to be methylated than CGI region, even in promoter regions. The distributions of methylation levels for CpG probes detected with *cis*-meQTLs are significantly different from the overall distributions. **(b)** Distribution of methylation levels in CGI, shelf and shore regions. **(c)** The distribution of proportion of explained variance (h^2) for 34,304 CpG probes with *cis*-meQTLs. **(d)** For *cis*-meQTLs, the explained phenotypic variances (h^2) depended on Δ, the distance between the peak SNP and target CpG-site. **(e)** The intensity plots of meQTLs for methylation-related genes based on EAGLE lung data. We detected *cis*-meQTLs for CpG probes located in three methylation related genes: *DNMT3A*, *DNMT3B* and *DNMT1*. P_E is the *P*-value based on EAGLE data and P_T is the P-value based on TCGA lung replication data. All of the associations were replicated in TCGA lung data. Note that one SNP, rs6058869, was associated with three probes in *DNMT3B*. The p-values were based on *t*-test. The box plots show the distribution of the methylation levels in each genotype category with error bars representing the 25% and 75% quantiles.

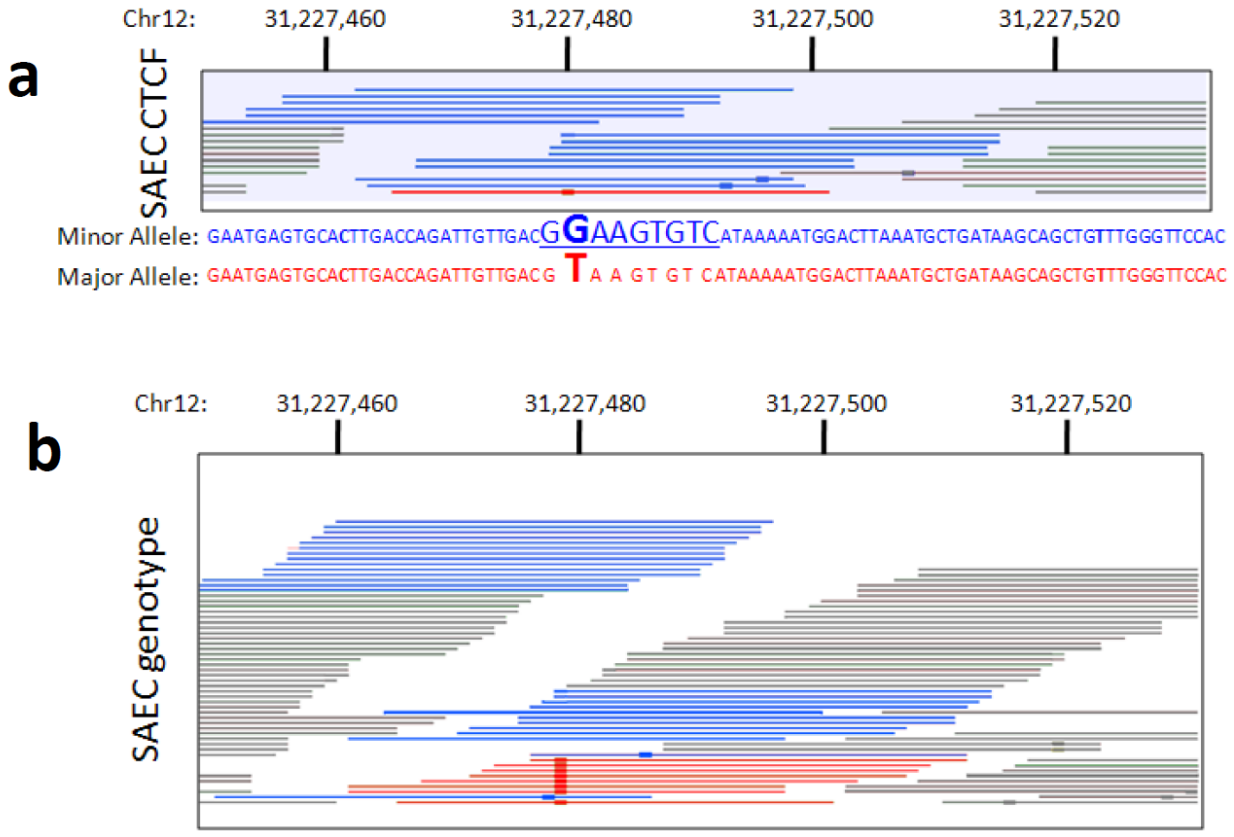


Supplementary Figure 2. The distribution of methylation levels for five CpG probes associated with the master regulatory SNP rs12933229.
The different distributions in three genotype categories in each plot show the gene association of the SNP with the methylation levels at the probe. Analysis was based on 210 EAGLE samples. The box plots show the distribution of the methylation levels in each genotype category with error bars representing the 25% and 75% quantiles.



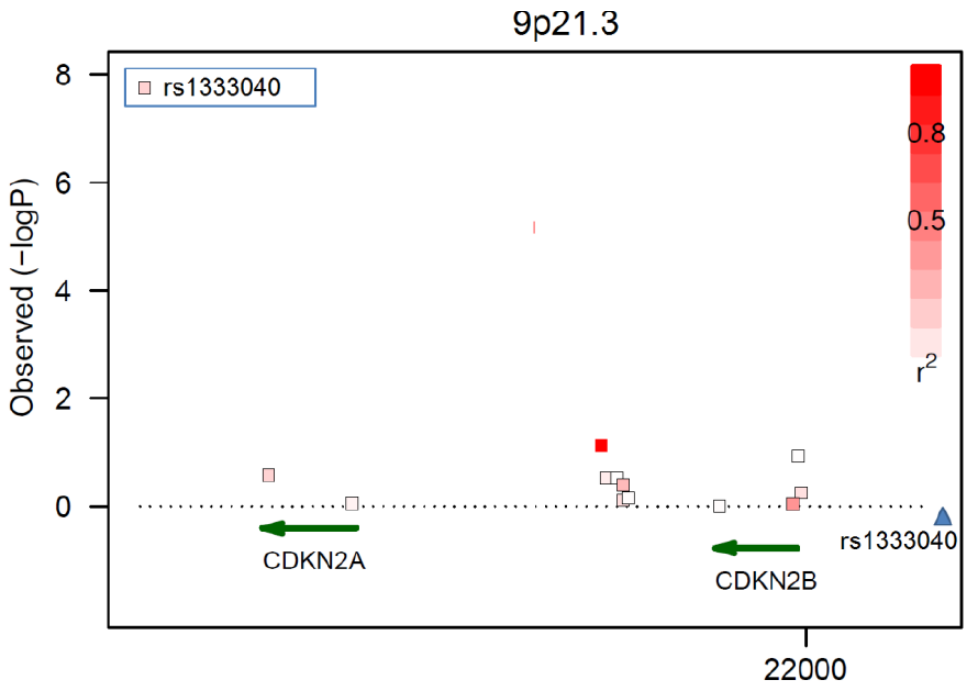
Supplementary Figure 3. A *cis*-meQTL SNP rs2816040 shows allele-specific CTCF ChIP.

CTCF ChIPseq data from ENCODE SAEC in the region containing rs2816057 (chr.1), in LD (r^2=0.98) with a *cis*-meQTL SNP rs2816040. When the minor C allele is present, CTCFBSDB 2.0 predicts a stronger CTCF binding site than when the major G allele is present. The observation of 86 'C' alleles v.s. 5 'G' allele suggests strong allele-specific binding at this site (P=2.0×10^{-20}, binomial test). Blue = Sequence reads containing minor allele SNP, red = sequence reads containing major allele SNP, gray = reads not containing SNP but in near proximity. Colored wide tick marks indicate deviation from the reference genome. The sequence at the bottom of the figure shows the precipitated region, with the minor allele sequence in colored blue and the predicted CTCF binding site underlined; the major allele that lacks the predicted site is in colored red.



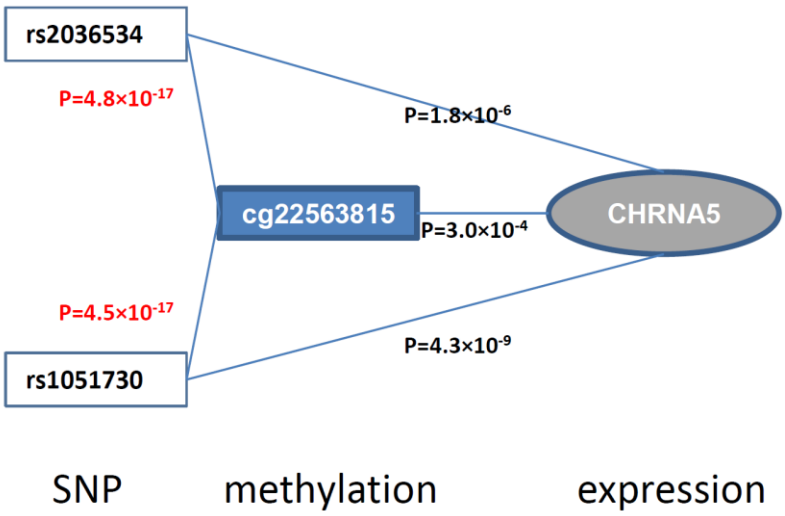
Supplementary Figure 4. A *cis*-meQTL SNP rs10843881 shows allele-specific CTCF ChIP.

CTCF ChIP-seq data from ENCODE SAEC in the region containing rs7305885 (chr.12), in LD (r^2=0.95) with a *cis*-meQTL SNP rs10843881, which affects a CpG probe in *cis* and a CpG probe in *trans*. When the minor G allele is present, CTCFBSDB 2.0 predicts a CTCF binding site. Figure **(a)**: The observation of 14 'G' alleles v.s. 1 'T' allele suggests strong allele-specific binding at this site (P=4.9×10^{-4}, bionomial test). Figure **(b)** shows all other publically available ENCODE-generated SAEC sequencing data except CTCF ChIP-seq, suggesting that the SNP is heterozygous in the cell line.



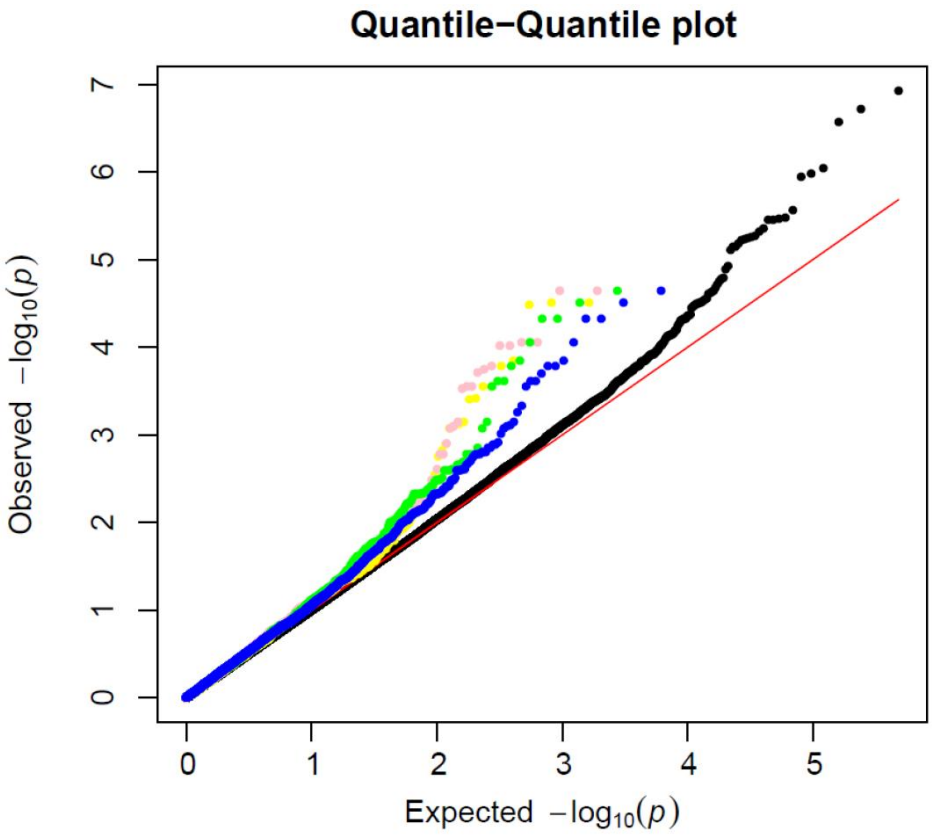
Supplementary Figure 5. Lung cancer GWAS risk SNP at 9p21.3 is not associated with the methylation of nearby CpG probes.

Methylation regional association for lung cancer GWAS SNP rs1333040 on chromosome 9p21.3. Symbols represent the association between the established lung cancer GWAS genetic locus in *CDKN2A* and methylation levels in nearby CpG-sites. Y-coordinate, *P*-value for association; x-coordinate, genomic location. For each SNP, the red solid square represents the methylation probe with the strongest association, whereas other methylation probes are colored on the basis of their correlation (measured as r^2) to the most-associated probe. The blue triangle points to the location of the risk SNP. No strong association was detected with nearby CpG probes at this locus, possibly because of low number of probes in this region on the Illumina platform.



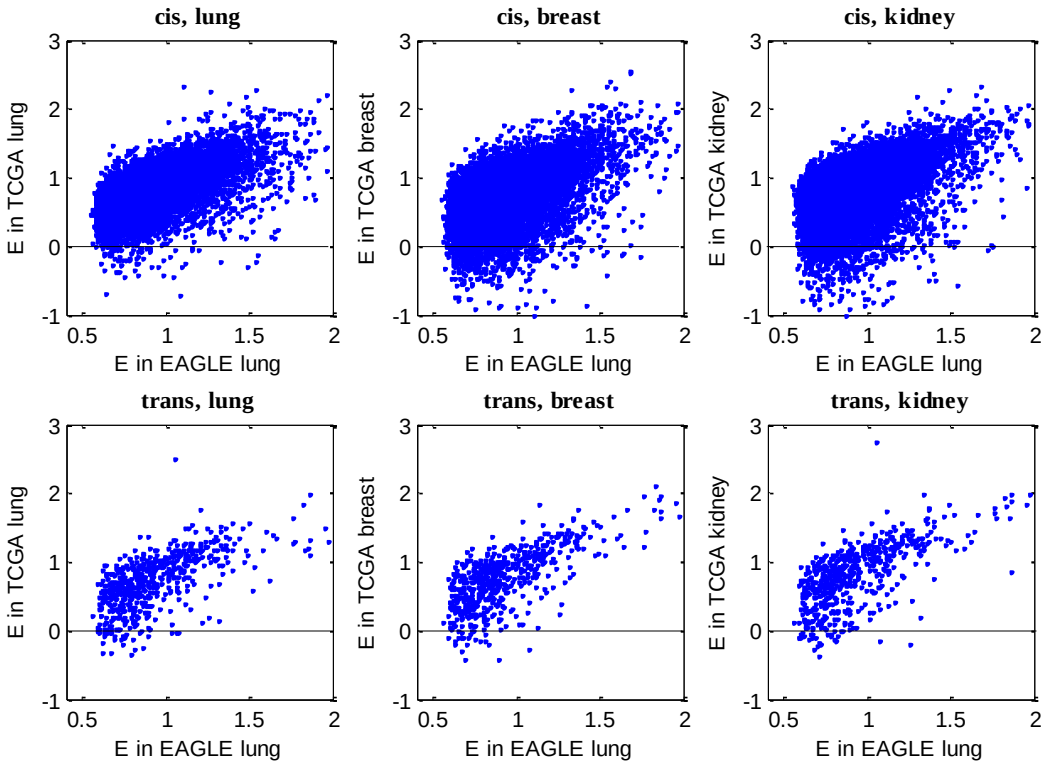
Supplementary Figure 6. Lung cancer GWAS risk SNPs at 15q25.1 are jointly associated with methylation and gene expressions.

Two SNPs in weak LD (r^2=0.17) at 15q25.1 are associated with lung cancer risk. We tested the genetic association between the two SNPs with the CpG probes (EAGLE n=210 samples) in the *CHRNA5* gene region and with the total expression of *CHRNA5* (TCGA n=59 samples, RNA-Seq). We also tested the association between the cg22563815 with *CHRNA5* expression (TCGA n=28 samples, methylation and RNA-Seq). The red *P*-values were based on EAGLE; the black *P*-values were based on TCGA data. We observed that jointly testing two SNPs substantially increased the significance for the association with both methylation of the CpG probe and total expression of *CHRNA5*. For total expression of *CHRNA5*, when analyzing two SNPs jointly, the *P*-value for rs2036564 was improved from 0.15 to 1.8×10^{-6} and for rs1051730 the *P*-value was improved from 0.0002 to 4.3×10^{-9}. For the methylation levels, joint analysis improved the *P*-value from 1.1×10^{-6} to 4.8×10^{-17} for rs2036534 and from 6.3×10^{-7} to 4.5×10^{-17} for rs1051730. Two SNPs jointly explained about 40% of the variation of methylation levels and 38% of the expression levels. The association between the SNPs and expression of *CHRNA5* was partially mediated by the CpG probe cg22563815. *P*-values were based *t*-test.



Supplementary Figure 7. Quantile-quantile (QQ) plots for associations between SNPs and the risk of developing lung squamous cell carcinoma (SQ).

Analyses based on 1424 cases and 4332 controls in NCI lung cancer GWAS. The SNPs in the established risk loci regions (15q25.1, 5p15.33, 6p21.33, 12p13.3 and 9p21.3) were excluded from analysis. Black: QQ plot for all autosomal SNPs. Genomic control (GC) λ-value is 1.00, suggesting no inflation. Blue: QQ plot for the meQTL SNPs associated with CpG probes mapping to north shore regions. λ-value is 1.12. Green: QQ plot for the meQTL SNPs associated with CpG probes mapping to north shore and gene body regions. λ-value is 1.14. Pink: QQ plot for the meQTL SNPs that overlap with CTCF binding regions and are associated with CpG probes mapping to non CGI and gene body regions. λ-value is 1.08. Yellow: QQ plot for the meQTL SNPs that overlap with H3k27me3 mark and are associated with CpG probes mapping to non CGI and gene body regions. λ-value is 1.05.



Supplementary Figure 8. Replication of EAGLE lung meQTLs in TCGA lung, breast and kidney normal tissue samples.

Each plot presents the estimated effect sizes E (based on linear regression) of *cis*-meQTLs in EAGLE lung discovery sample (x-coordinate) v.s. E in TCGA replication samples (y-coordinate). These effect sizes are comparable across studies because all traits were quantile-normalized to follow the standard normal distribution. The dots above the horizontal line show consistent directions in discovery and replication samples. The plots are based on all 34,304 *cis*-meQTLs and 585 *trans*-meQTLs.

Supplementary Table 1: Number of CpG probes, *cis*-meQTLs and proportion of probes detected with *cis*-meQTLs.

	Number of CpG probes passing QC			Number of CpG probes detected with *cis*-meQTLs			Proportion of CpG probes detected with *cis*-meQTLs		
	CGI[a]	non-CGI[b]	Combined[c]	CGI	non-CGI	combined	CGI	non-CGI	combined
TSS1500	23504	35859	59363	1354	4517	5871	5.8%	12.6%	9.9%
TSS200	31498	16626	48124	1520	1853	3373	4.8%	11.1%	7.0%
5'UTR	25414	23194	48608	1212	2462	3674	4.8%	10.6%	7.6%
1st Exon	21884	9327	31211	866	940	1806	4.0%	10.1%	5.8%
Body	35146	89689	124835	3322	9242	12564	9.5%	10.3%	10.1%
3'UTR	2009	12701	14710	221	1046	1267	11.0%	8.2%	8.6%
	gene[d]	non-gene[e]	combined[f]	gene	non-gene	combined	gene	non-gene	combined
N_Shelf	9818	4658	14476	899	528	1427	9.2%	11.3%	9.9%
N_Shore	35853	8285	44138	4219	1193	5412	11.8%	14.4%	12.3%
CGI	103990	17804	121794	6895	2603	9498	6.6%	14.6%	7.8%
S_Shore	28645	5869	34514	3484	922	4406	12.2%	15.7%	12.8%
S_Shelf	8546	4329	12875	744	444	1188	8.7%	10.3%	9.2%
All[g]			338456			34304			10.1%

[a]: number of CpG probes located in CpG islands.
[b]: number of CpG probes not located in CpG islands.
[c]: CGI + non-CGI
[d]: number of probes annotated in genes
[e]: number of probes annotated not in genes
[f]: genes + non-genes.
[g]: analysis based on all CpG probes passing QC filters.

Supplementary Table 2: Enrichment analysis of genes harboring CpG probes mediating *trans*-associations.

Term	Count	%	*P*-Value	Genes	Fold Enrichment
GO:0043547~positive regulation of GTPase activity	3	2.7	0.0037	SH3D20, TSC2, ARHGAP27	32.4
GO:0051056~regulation of small GTPase mediated signal transduction	6	5.3	0.0038	TBC1D14, SH3D20, TSC2, RASGEF1B, ARHGAP27, RASA3	5.7
GO:0043087~regulation of GTPase activity	4	3.5	0.014	TBC1D14, SH3D20, TSC2, ARHGAP27	7.7
GO:0043089~positive regulation of Cdc42 GTPase activity	2	1.8	0.021	SH3D20, ARHGAP27	94.9
GO:0006350~transcription	16	14.2	0.021	ZNF430, ZNF273, ZNF100, ZFP57, DEDD, BANP, HES6, MED7, PCGF3, ZNF429, NCOA4, HEY2, ZNF107, ZNF431, ZNF493, ZNF257	1.8
GO:0006349~genetic imprinting	2	1.8	0.041	ZFP57, KCNQ1	47.5
GO:0032489~regulation of Cdc42 protein signal transduction	2	1.8	0.045	SH3D20, ARHGAP27	43.2
GO:0043088~regulation of Cdc42 GTPase activity	2	1.8	0.045	SH3D20, ARHGAP27	43.2
GO:0051186~cofactor metabolic process	4	3.5	0.047	BLVRA, SDHA, CTNS, NAPRT1	4.9

DAVID analysis based on Gene Ontology classification of genes in the region of the proximal CpG sites mediating the *trans*-associations. *P*-values were based on Fisher Exact test.

Supplementary Table 3: Master regulatory SNP rs12933229 and its associated distal CpG probes across tissue types.

					EAGLE lung		TCGA lung		TCGA breast		TCGA kidney	
CpG probe	chr	base pair	gene	gene location[a]	β	*P*	β	*P*	β	*P*	β	*P*
cg01325958	17	37793326	STARD3	TSS200	0.81	5.8E-16	0.57	2.5E-03	0.81	3.4E-7	0.99	4.9E-16
cg02201215	14	105714846	BRF1	Body TSS200, TSS1500	0.86	5.4E-18	0.74	5.0E-05	0.74	2.5E-6	0.78	1.7E-9
cg11143796	4	178363279	AGA	Body	0.77	3.7E-15	0.61	3.8E-04	0.83	1.5E-7	0.99	4.6E-16
cg16383001	1	100435543	SLC35A3	5'UTR 1stExon	0.69	3.2E-11	0.51	3.3E-03	0.75	2.7E-6	0.84	3.6E-11
cg16510040	1	40042432	PABPC4	1stExon 5'UTR	0.70	4.3E-11	0.31	4.0E-02	0.48	0.004	0.60	8.0E-6

[a]:Gene location may vary based on different transcripts.

SNP rs12933229 was associated with five CpG probes on different chromosomes. The associations were replicated in TCGA lung, breast and kidney histologically normal tissue samples. All CpG probes are in CpG islands. The reference allele is "C". β is the coefficient in the linear regression and *P* is the two-sided p-value.

Supplementary Table 4: Enrichment analysis of genes that harbor *cis*-meQTLs in north shore and gene body regions.

Term	n	P-Value	Genes	Fold Enrichment
hsa04144:Endocytosis	27	2.0E-04	PRKCZ, ERBB4, AP2S1, PIP5K1C, ASAP3, IGF1R, SH3GLB2, SNF8, DNAJC6, NEDD4L, AGAP1, IQSEC1, EHD4, RET, RUFY1, HLA-C, HLA-B, KDR, RAB11FIP5, AP2A2, AP2A1, NEDD4, NTRK1, ACAP1, RAB22A, GRK7, PARD6G, ARAP1	2.18
hsa05200:Pathways in cancer	40	2.5E-04	FGF6, DCC, FGF18, WNT5B, ERBB2, NFKBIA, FOXO1, GLI2, SHH, AKT1, IGF1R, MAX, CASP9, SOS1, PAX8, RAC1, AXIN1, PIK3R2, FN1, RET, MSH3, MAP2K2, RXRB, FLT3, RXRA, FGF22, RB1, FGF20, DAPK3, RALGDS, CDK2, CTNNA2, MAPK1, RASSF5, ETS1, LAMA5, NTRK1, WNT11, WNT9A, WNT7A	1.81
hsa00062:Fatty acid elongation in mitochondria	5	1.1E-03	PPT2, ECHS1, PPT1, HADHA, HADHB	9.29
hsa05223:NSCLC	11	2.7E-03	AKT1, MAPK1, RASSF5, CASP9, MAP2K2, RXRB, ERBB2, SOS1, RXRA, RB1, PIK3R2	3.03
hsa04514:Cell adhesion molecules (CAMs)	18	6.6E-03	HLA-DQB1, F11R, CLDN18, NRXN2, HLA-DRB1, NRXN3, HLA-C, CLDN10, HLA-B, SDC4, CDH4, ALCAM, CD34, CLDN1, HLA-DRB5, CNTNAP1, HLA-DPB1, HLA-DOA, JAM3	2.03
hsa05216:Thyroid cancer	7	1.1E-02	MAPK1, RET, MAP2K2, RXRB, NTRK1, RXRA, PAX8	3.59
hsa04360:Axon guidance	17	1.2E-02	DCC, NGEF, PLXNC1, PLXNB2, EPHB3, NTN1, SLIT1, MAPK1, EPHA4, SEMA4G, RAC1, ROBO2, EFNA5, NFATC4, ROBO3, UNC5C, NFATC1	1.96
hsa05215:Prostate cancer	13	1.5E-02	MAP2K2, ERBB2, FOXO1, NFKBIA, RB1, CDK2, INSRR, AKT1, IGF1R, MAPK1, CASP9, SOS1, PIK3R2	2.17
hsa04010:MAPK signaling pathway	28	1.9E-02	FGF6, FGF18, GNA12, MKNK1, DAXX, AKT1, MAX, BDNF, MAP3K5, MAPT, SOS1, RAC1, NFATC4, CACNG8, MAP2K2, FGF22, FGF20, CACNA2D4, MAPK1, NTRK1, MAPK8IP2, CACNA1H, MAPK8IP1, DUSP8, CACNA1A, CACNA1B, MAP3K11, DUSP6	1.56
hsa04940:Type I diabetes mellitus	8	2.0E-02	HLA-DQB1, PRF1, HLA-DRB1, PTPRN2, HLA-DRB5, HLA-C, HLA-B, HLA-DPB1, HLA-DOA	2.83
hsa05213:Endometrial cancer	9	2.1E-02	AKT1, MAPK1, CASP9, MAP2K2, ERBB2, SOS1, AXIN1, PIK3R2, CTNNA2	2.57
hsa05222:Small cell lung cancer	12	2.4E-02	AKT1, MAX, CASP9, LAMA5, RXRB, RXRA, NFKBIA, RB1, APAF1, CDK2, PIK3R2, FN1	2.12

We found that SNPs associated with CpG probes located in gene bodies and north shores were enriched for squamous cell carcinoma risk. We examined whether genes harboring these CpGs are enriched in cancer related pathways or gene sets by performing Gene Ontology analysis. The table lists the 12 most enriched gene sets or KEGG pathways. Interestingly, they are largely enriched for cancer-related genes. *P*-values were based on Fisher Exact test.