**Overcoming micro-environmental stresses through DNA methylation in cancer**

**Introduction**

**Materials and Methods**

**Penalized linear regression models**

For each gene expression level, we selected its proximal CpG islands, and/or regulating transcription factors’ expressions.

As the generated models have different numbers of predictors for each gene, we picked best-performing models with 2 to 5 predictors in each GLM so that the prediction powers could be comparable.

**Isoform and DNA methylation data**

From TCGA, we collected XXX sets of isoform gene expression data, and XXX sets of paired DNA methylation data.

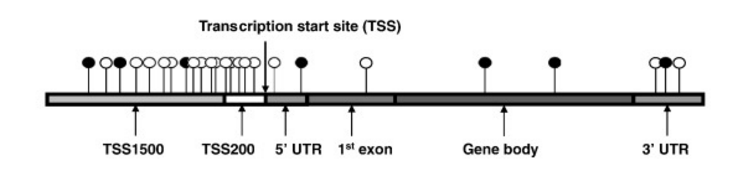
**Transcription factors and enhancers database**

From TCGA, we collected XXX sets of isoform gene expression data, and XXX sets of paired DNA methylation data. We collected all the transcription factors and their target genes from the following databases: TRED, Neph2012, ENCODE, Marbach2016, TRRUST; and enhancers and their target genes from Dendb.

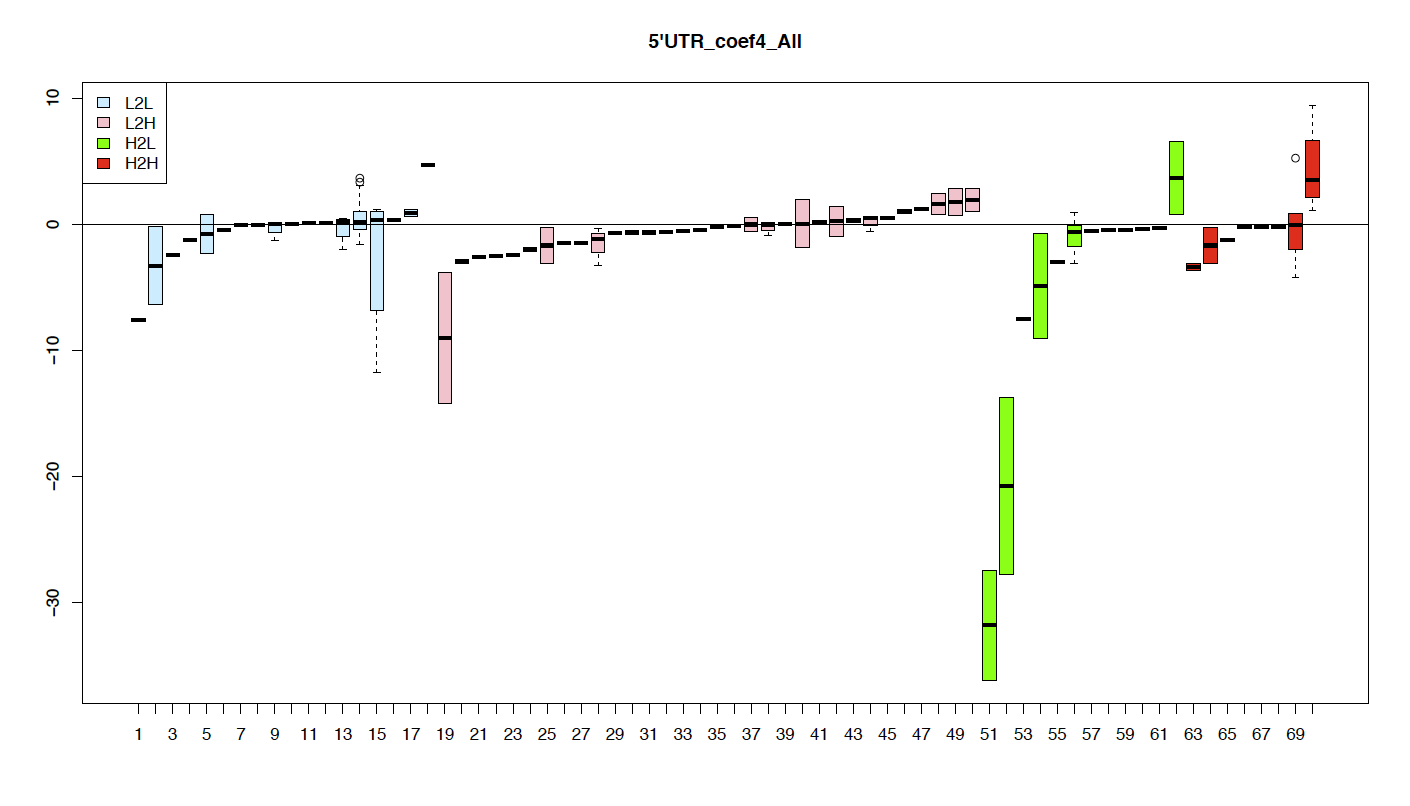
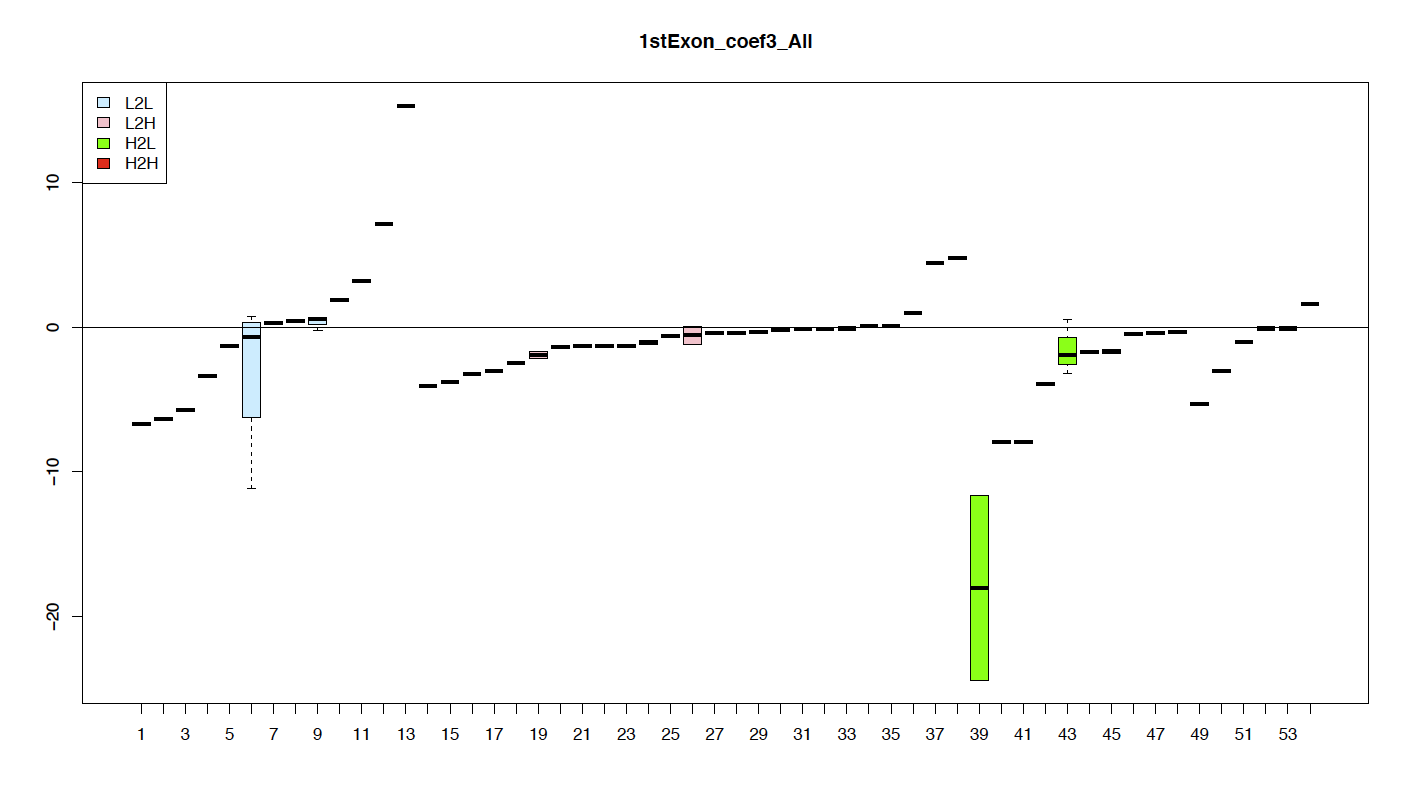
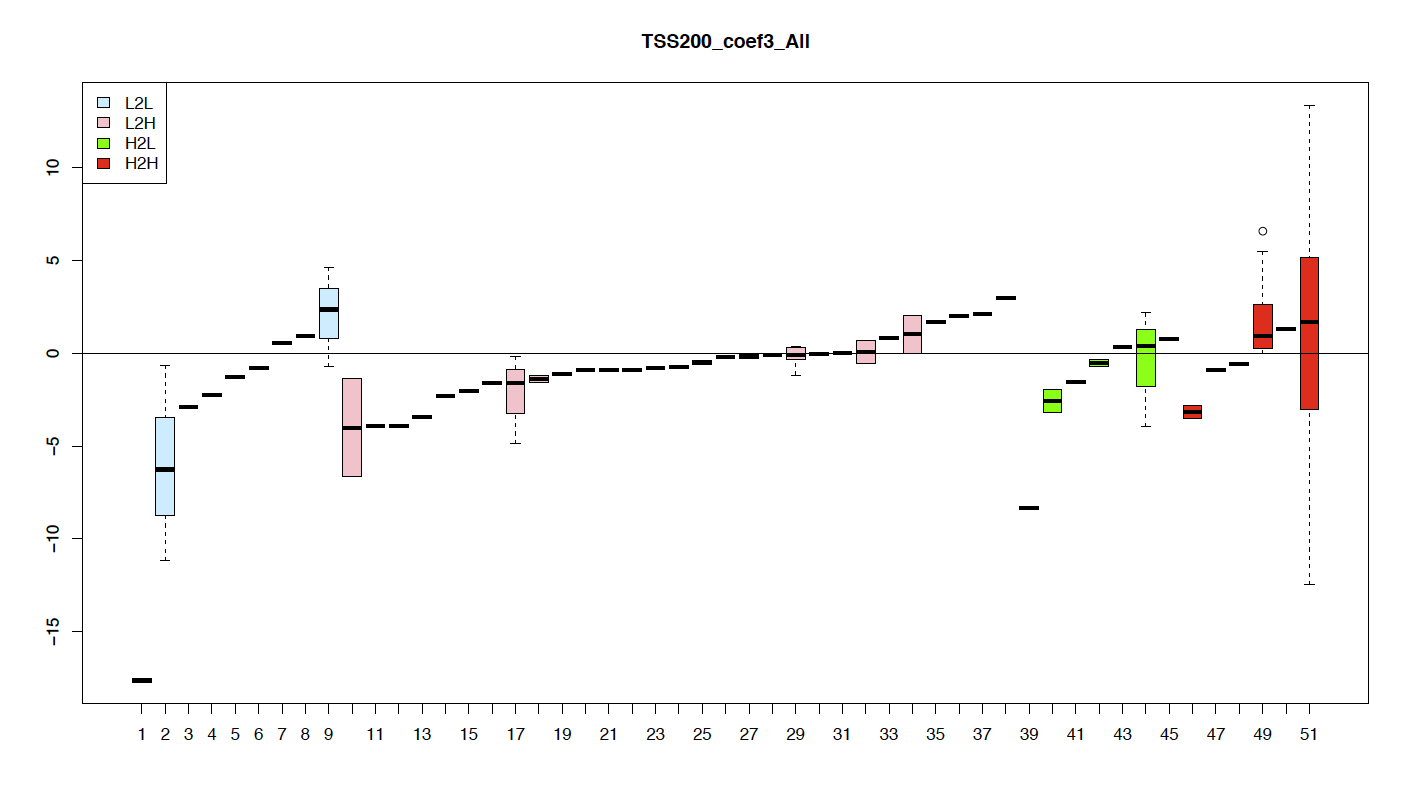
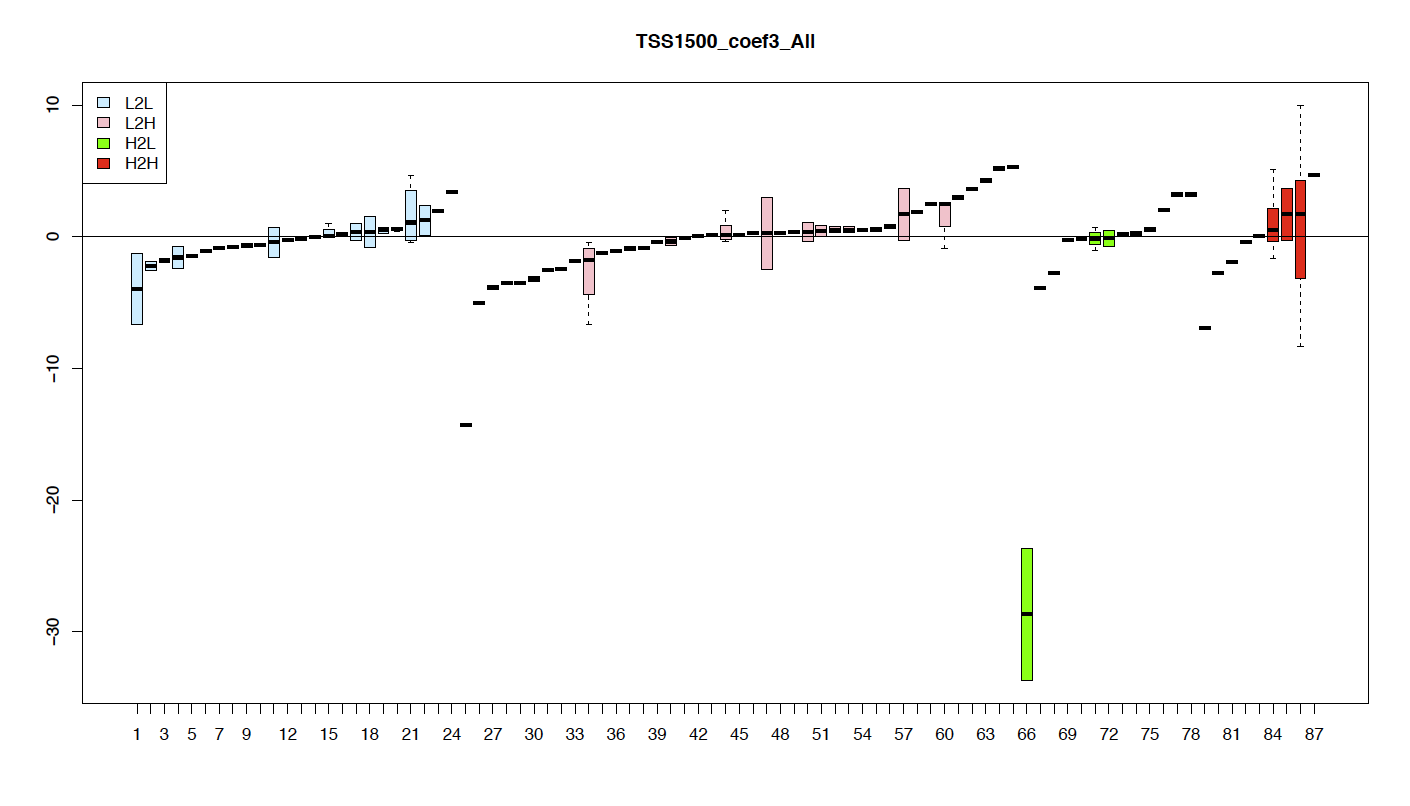
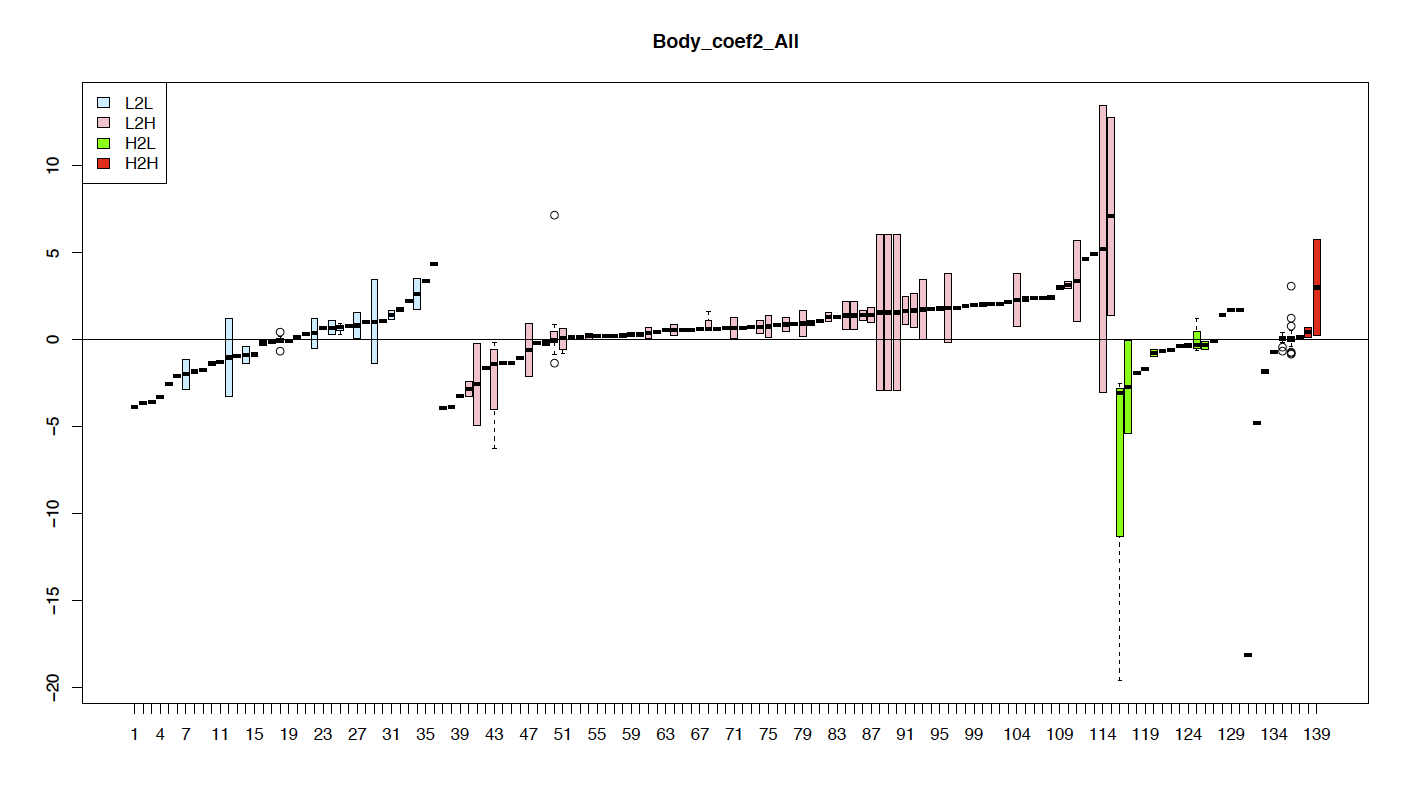
**Results**

**DNA methylation utilized different ways of regulations to help regulate the gene expressions**

In order for genes’ expression levels to elevate, they displayed very different changes in DNA methylation levels. For each gene, we took out a subset of its proximal CpG islands (CGIs) whose combination could predict the expression level of the gene, denoted as predictor CGIs. We further observed the coefficients of these predictor CGIs methylation levels. We categorized all the CGIs into six categories, including, TSS1500, TSS200, 5’UTR, 1st exon, gene body and 3’UTR, and noticed that for those genes to go from low expression to high expression (pink boxplots in Figure 2), CGIs contribute differently according to their relative locations on the genome: TSS200 CGIs goes down significantly, the same as TSS1500, but not as significant as TSS200, probably because TSS1500 CGIs are further away from the TSS, which means that they may have less impact on regulating the gene expressions. In the gene body region, exons CGIs need to go down, while majority of the gene body, which consists of introns, needs to go up. 5’UTR needs to go down, while on the contrary, 3’UTR CGIs needs to go up. We suspect XXX.



**Figure 1**: The relative locations of CGIs on the genome.



**Figure 2**: The distribution of coefficients of CGIs falling into TSS1500, TSS200, 5’UTR, 1st exon, gene body and 3’UTR, for genes whose expression values are changed from low to low (light blue), low to high (pink), high to low (green) and high to high (red).

The expression level changes of genes are subject to different patterns of DNA methylational changes in their proximal CpG islands.

**Genes’ whose expressions are regulated by epigenetic markers are enriched in micro-environmental stress responses**

We selected those genes whose expression levels could be well explained by its proximal CpG islands methylation levels in all tumor types, but not in any of the normal tissue types, and found out that there are 20 such genes. Their functions are all related to extracellular activities. We further loosen our criteria, and selected those genes whose expression levels could be well explained by methylation levels in most tumor types, but not in the normal tissue types, and a simple pathway enrichment analyses showed that these genes are enriched in extracellular matrix proteins.

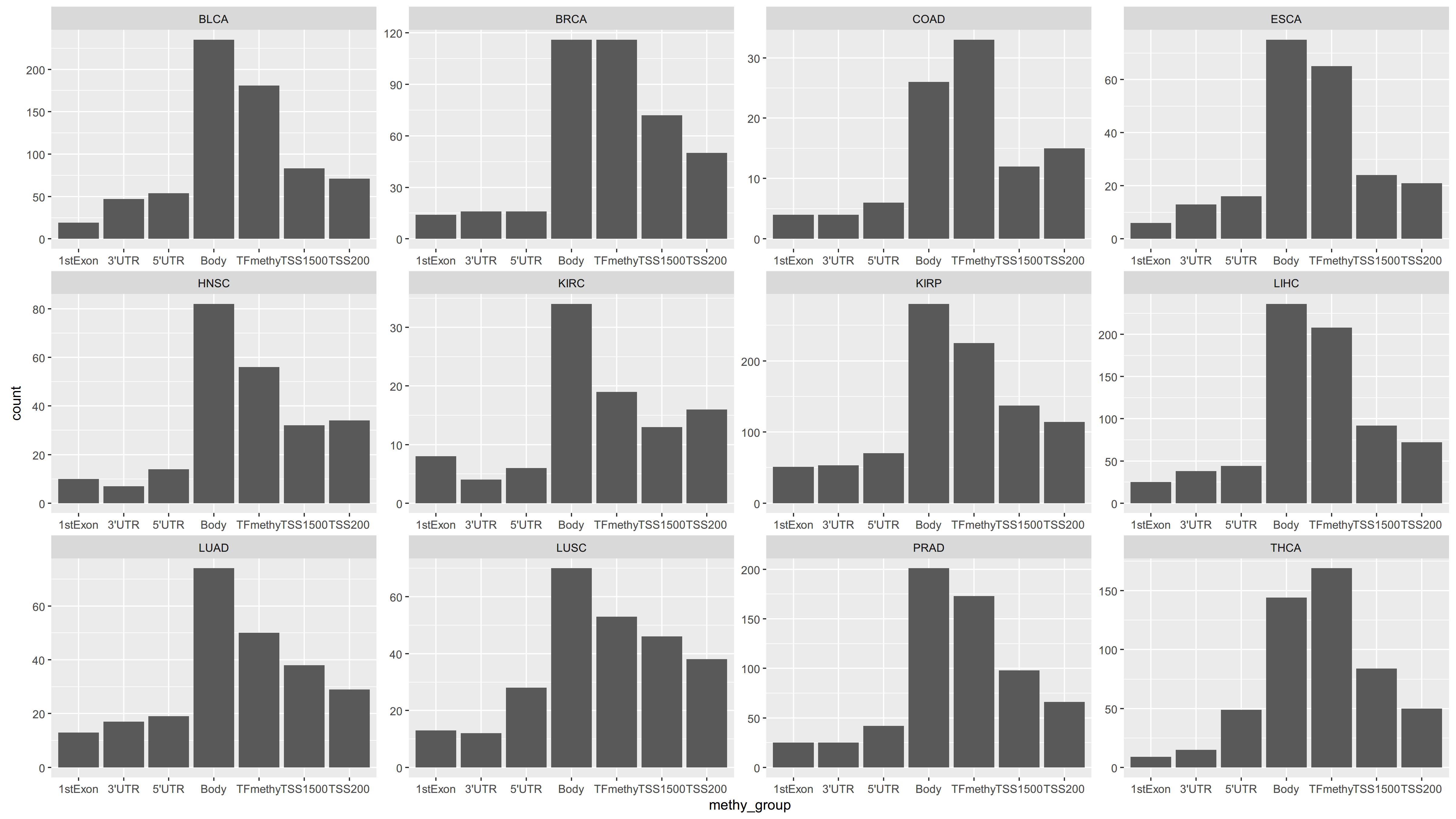
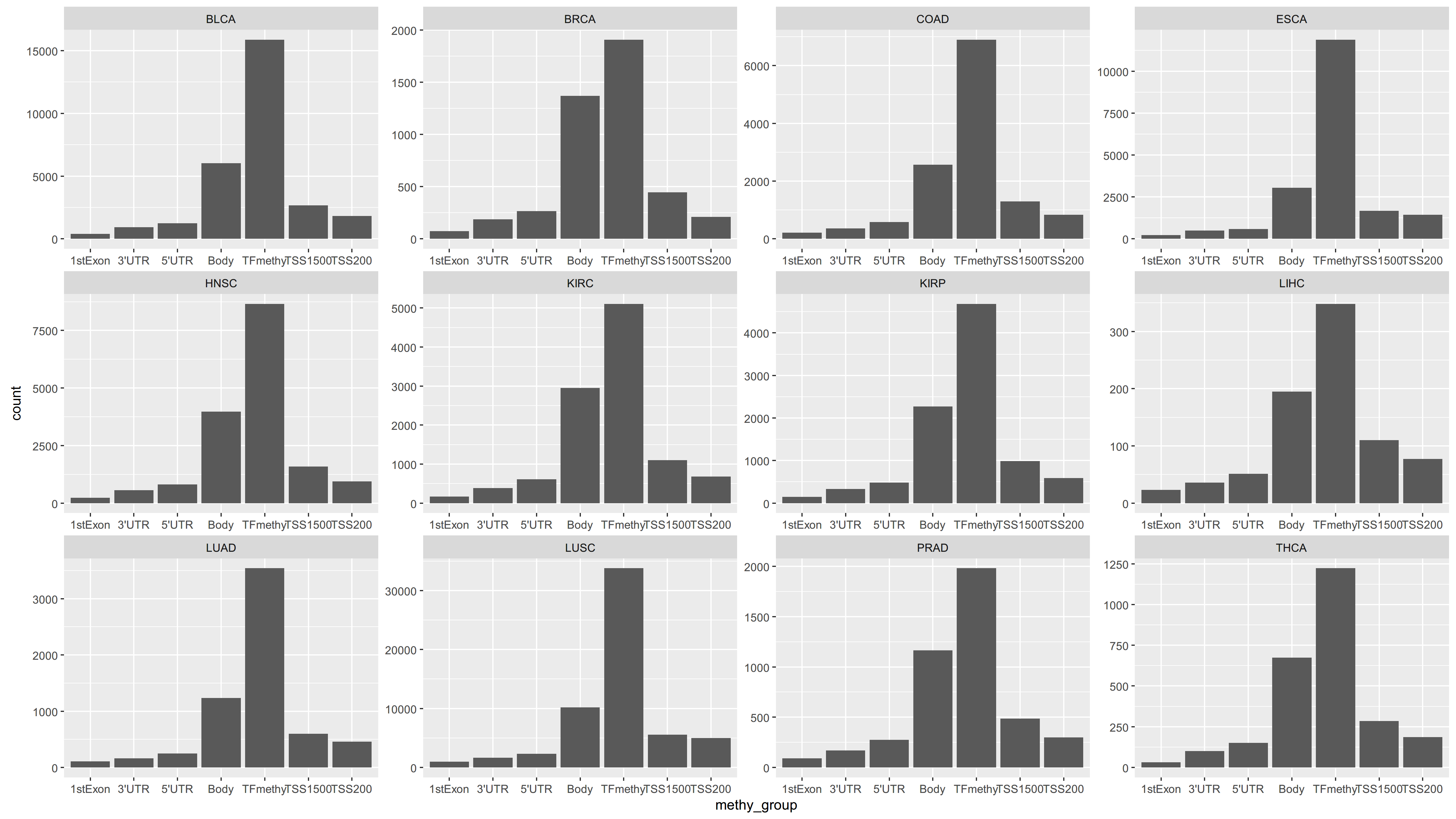
[Need one enrichment table]

**Gene expression changes are subject to transcriptional and methylational regulations in a complementary way**

Cancer cells live in a very unfriendly environment. They have been under tremendous pressure to evolve to overcome stress types resulted from various micro-environmental changes. In order to survive, they have to acquire various abnormal functionalities for them to by-pass or go undetected layers of defense mechanisms in the human body that are in place to prevent abnormalities from happening. Such new capabilities may be gained through adaption initially via functional regulation and then gradually via epigenomic and genomic changes selected to make their survival more sustainable and possibly more efficient. As the evolving cells accumulate increasingly more abnormalities, they also have to evolve to gain additional protective capabilities so they can survive the increasingly hostile environments as they become more abnormal. In order to see whether tumor cells utilized more epigenetic regulations, we built penalized linear model for each gene, using predictors of its proximal CGIs and those proximal CGIs of the gene’s regulating TFs. In normal cells, we noticed that, TF associated CGIs are predominantly the predictors for genes’ expressions; while in tumor cells, CGIs in the gene’s gene body regions stand out to be the prominent predictors.

A. Normal samples

B. Tumor samples



[The figure needs to be replaced by better quality.]

**Figure 3**: the prediction powers

**Discussion**

**Acknowledgements**

**References**