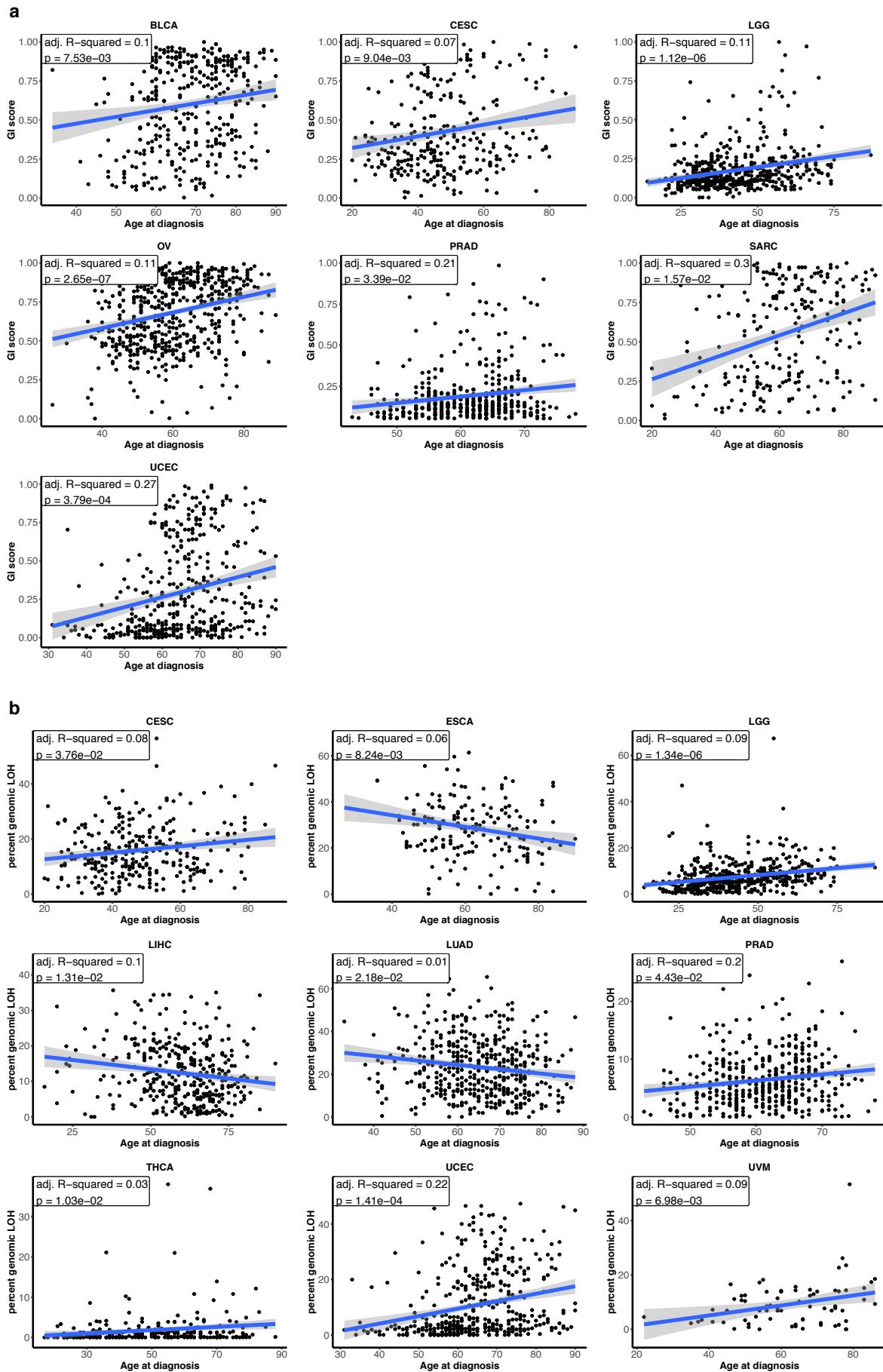


## **Supplementary Figures**

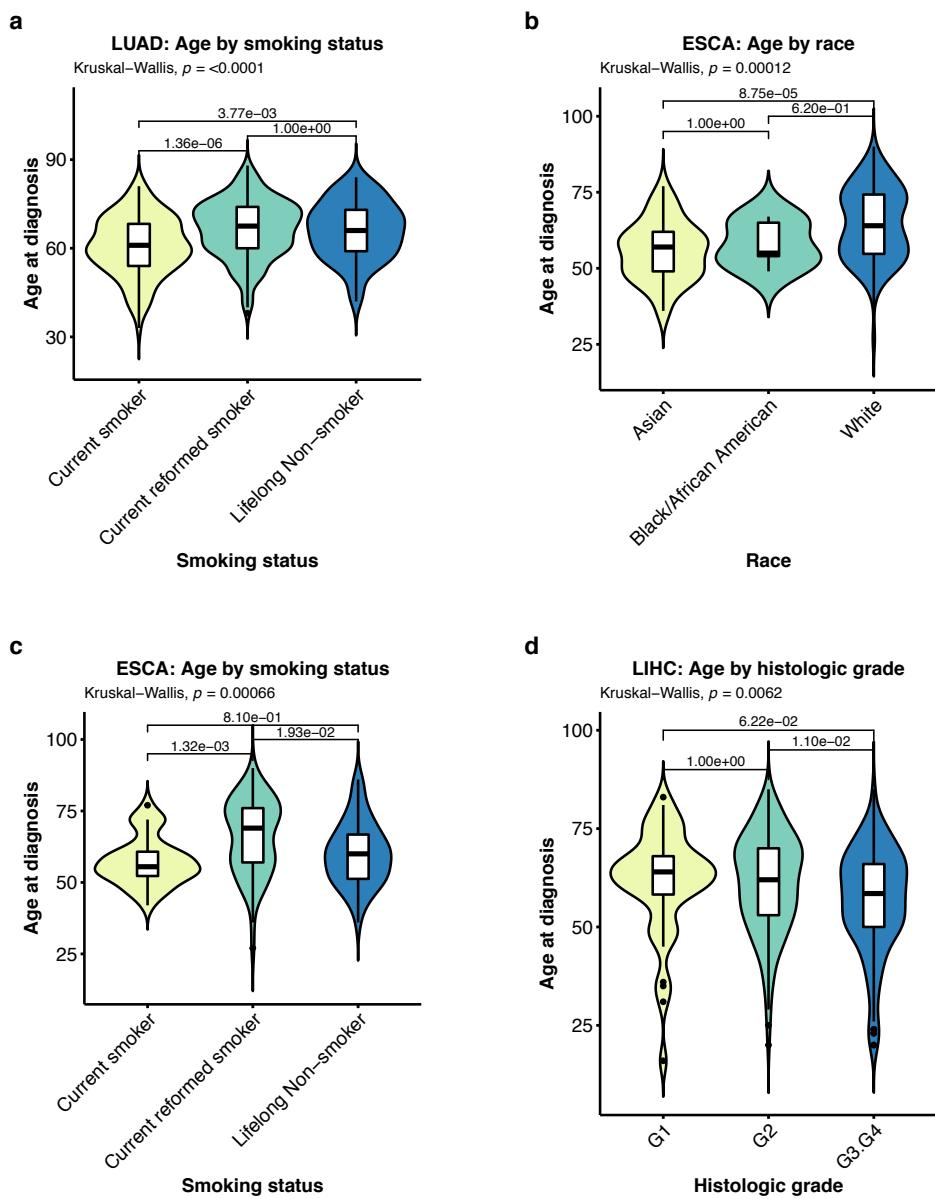
# **An Integrative Analysis of the Age-Associated Genomic, Transcriptomic and Epigenetic Landscape across Cancers**

Chatsirisupachai et al.



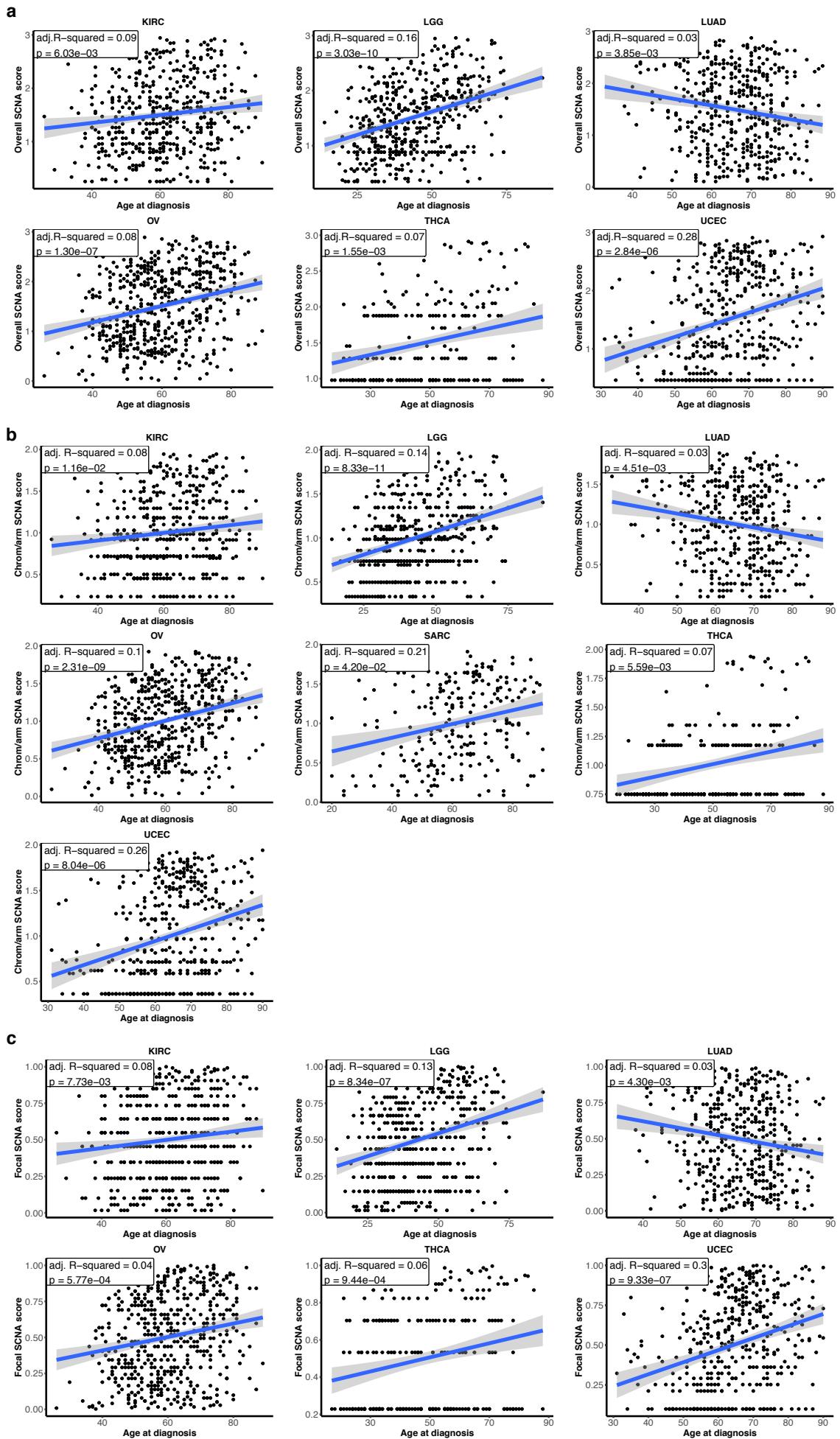
### **Supplementary Fig. 1**

Association between age and **(a)** GI score and **(b)** percent genomic LOH. Multiple linear regression was performed to identify the relationship between age and GI score or percent genomic LOH for each cancer type. Cancer types with a significant association (adj. p-value < 0.05) are shown together with adjusted R-squared and p-values from multiple linear regression analysis.



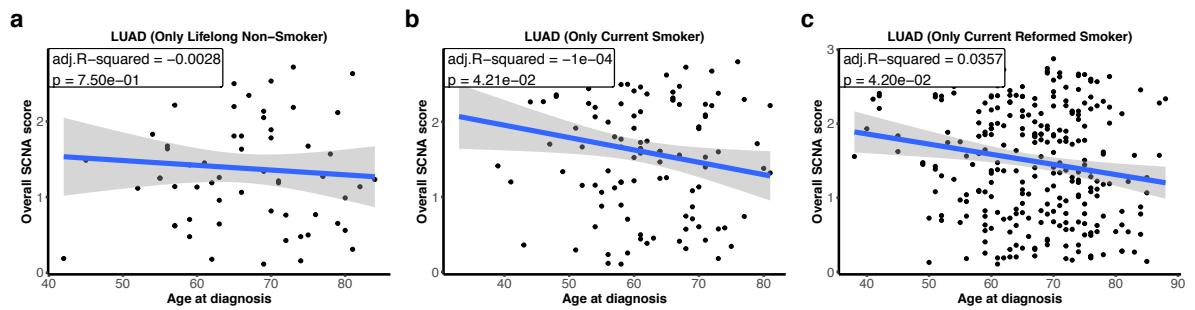
## Supplementary Fig. 2

Age distribution of patients presented with different clinical factors. **a** Age distribution of lung adenocarcinoma patients by smoking status. **b** Age distribution of oesophageal cancer patients by race. **c** Age distribution of oesophageal cancer patients by smoking status. **d** Age distribution of liver cancer patients by histologic grade. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were done using two-sided Dunn's test. The middle bar of the boxplot is the median. The box represents interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of  $1.5 \times$  IQR.



### **Supplementary Fig. 3**

Association between age and **(a)** overall SCNA score, **(b)** chromosome/arm-level SCNA score, and **(c)** focal-level SCNA score. Multiple linear regression was performed to identify the relationship between age and SCNA score for each cancer type. Cancer types with a significant association (adj. p-value < 0.05) are shown together with adjusted R-squared and p-value from multiple linear regression analysis.



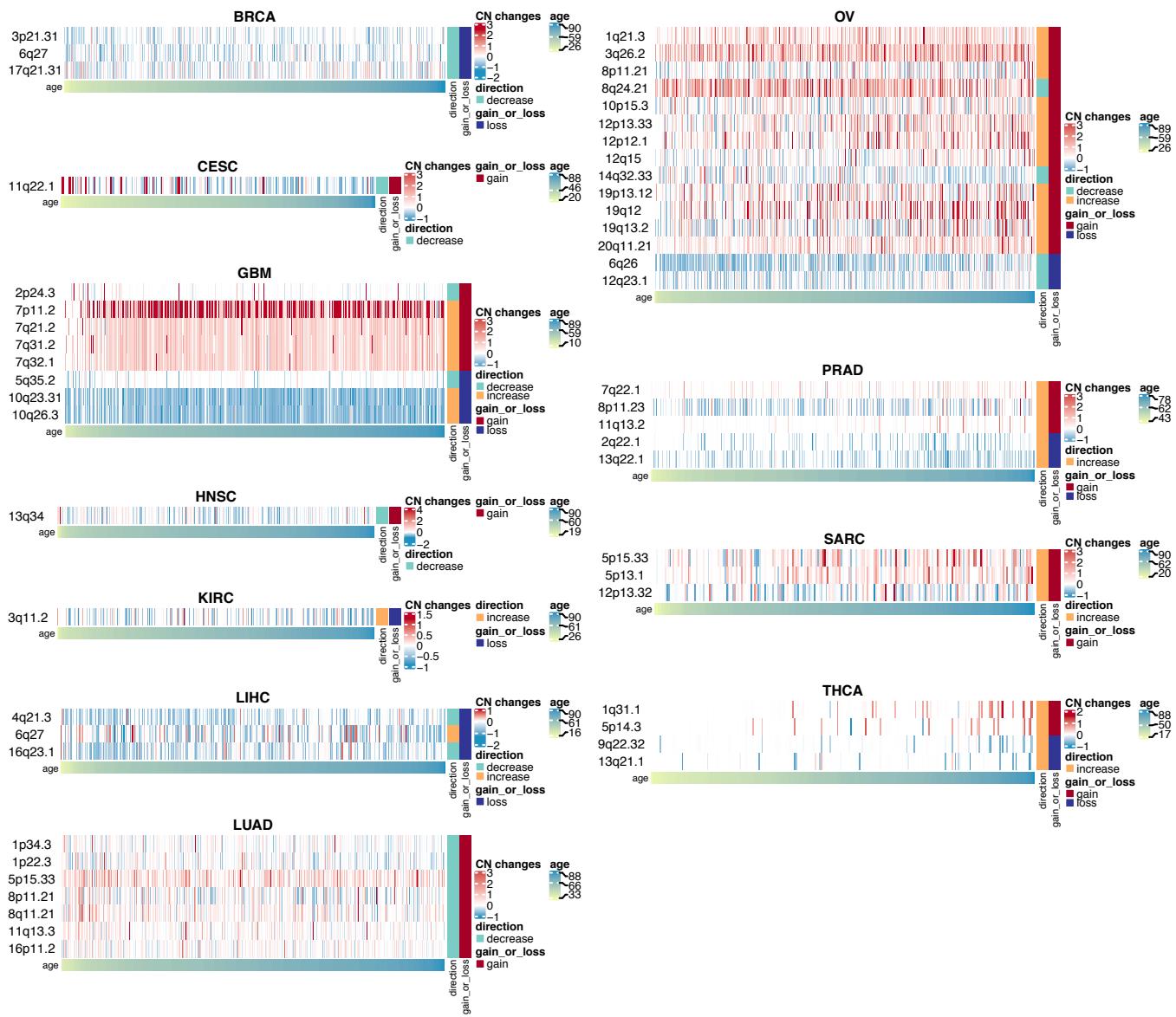
**Supplementary Fig. 4**

Association between age and overall SCNA score in lung adenocarcinoma separated by smoking status - (a) lifelong non-smoker, (b) current smoker and (c) current reformed smoker. Adjusted R-squared and p-value from multiple linear regression analysis are presented.



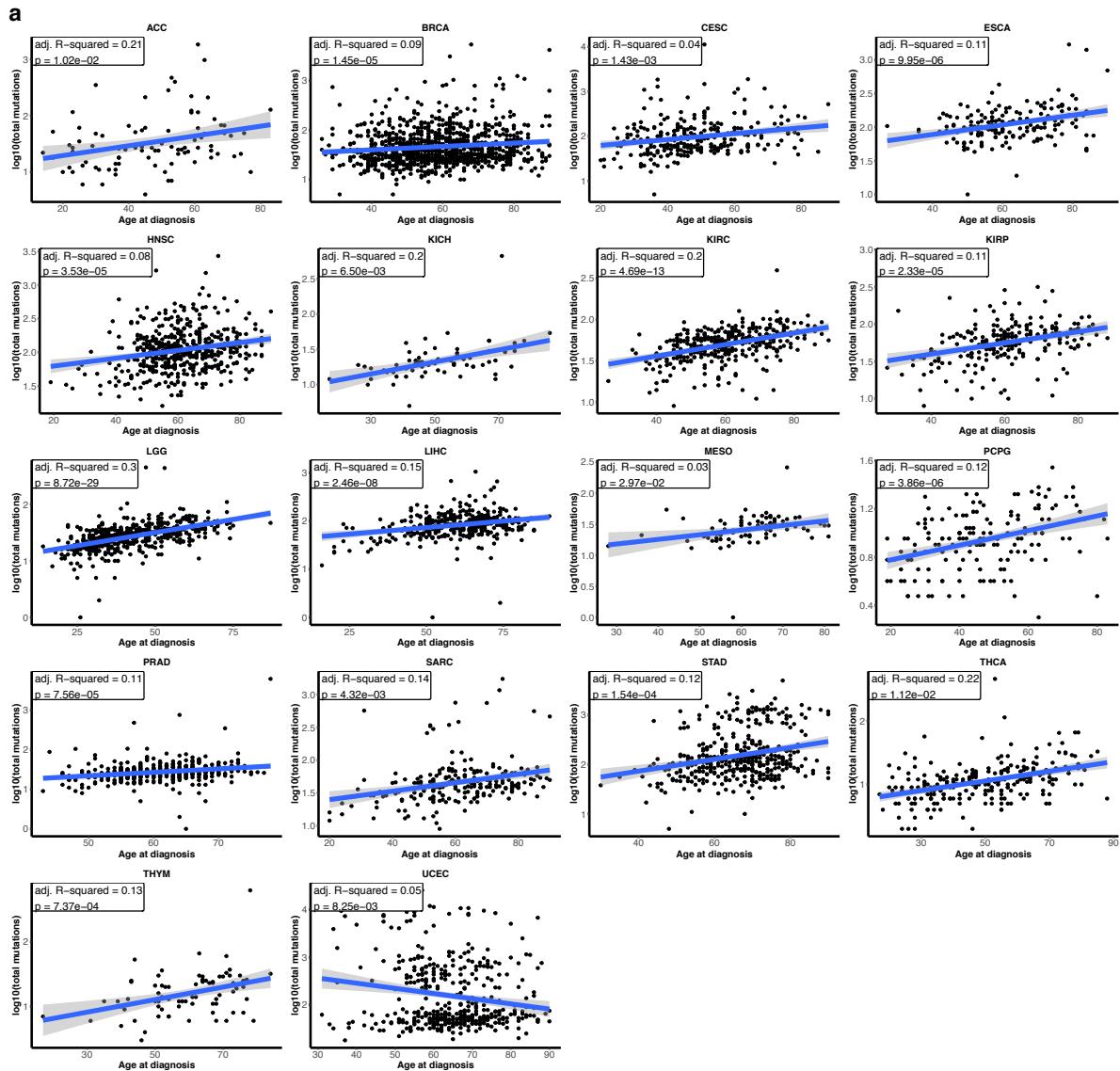
## Supplementary Fig. 5

Heatmaps represent arm-level copy-number alterations across cancer types. Samples are sorted by age. Colours represent copy-number changes from GISTIC2.0, blue denotes loss and red corresponds to gain.

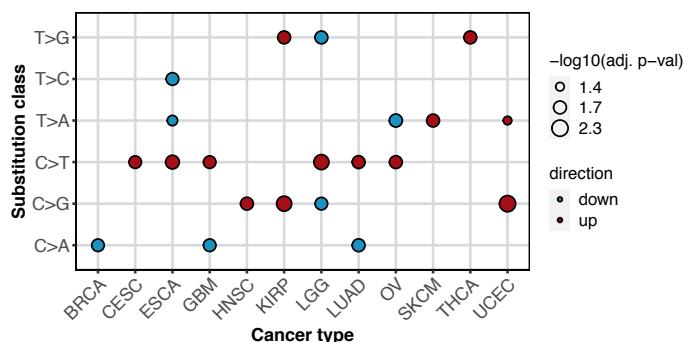


## Supplementary Fig. 6

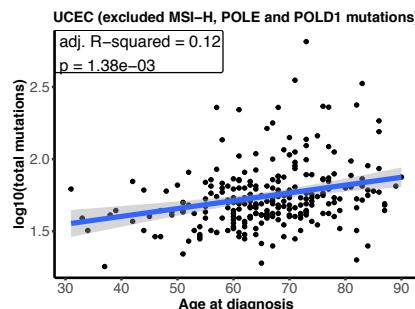
Heatmaps represent recurrent gain and loss of focal-regions that showed the age-associated patterns across cancer types. Samples are sorted by age. Colours represent copy-number changes from GISTIC2.0, blue denotes loss and red corresponds to gain. The direction legend shows whether the gain/loss of the region increases or decreases with age.



**b**

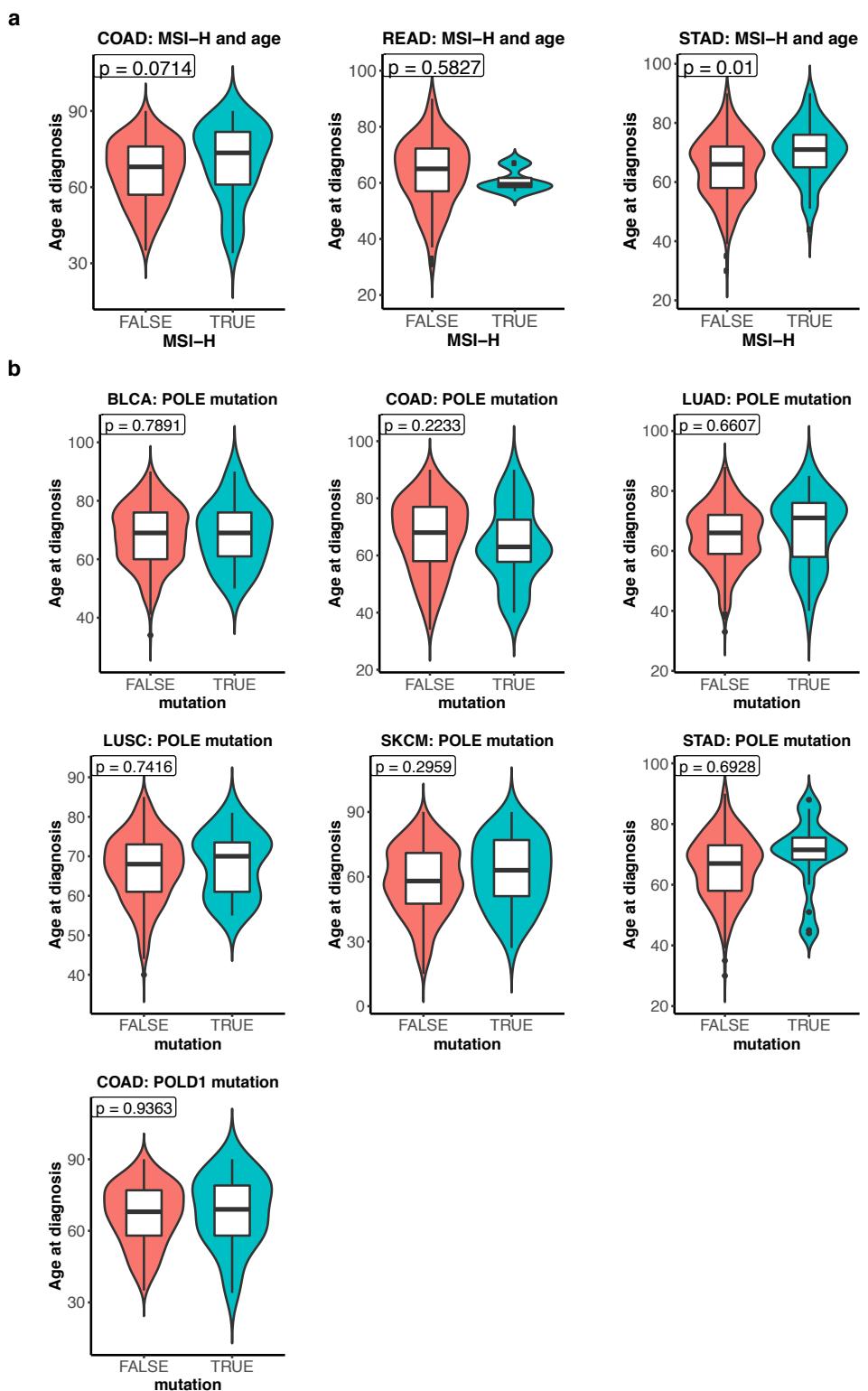


**c**



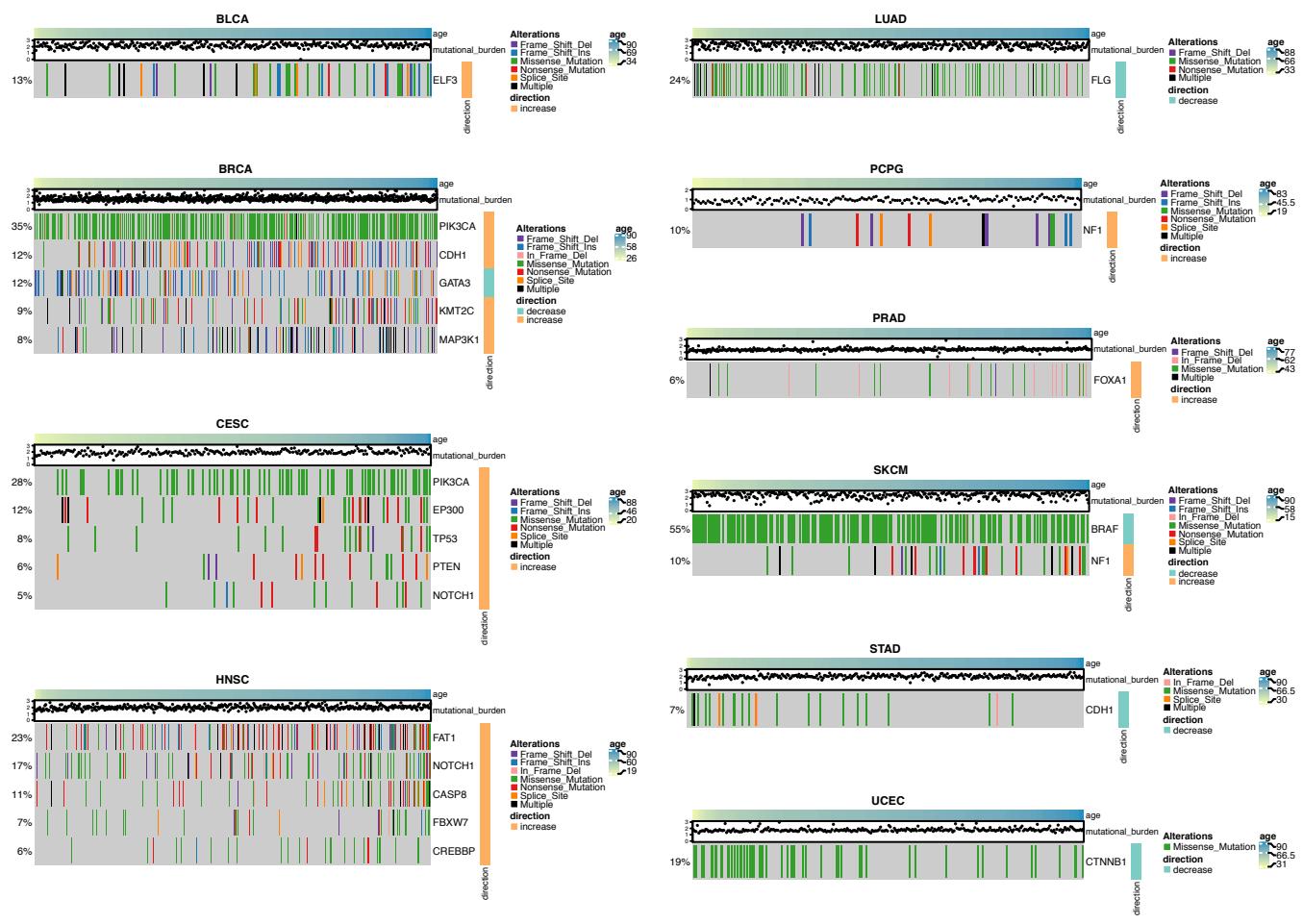
### **Supplementary Fig. 7**

Association between age and somatic mutations. **a** Association between age and mutational burden. Simple linear regression was performed to investigate the association between age and mutational burden. Cancer types with a significant association (adj. p-value < 0.05) were further investigated using multiple linear regression. The figure shows cancer type-specific analysis for every cancer that had a significant association in the simple linear regression analysis. Adjusted R-squared and p-value from multiple linear regression analysis are displayed. **b** Association between age and fraction contribution of each substitution class. For each substitution class in each cancer, simple linear regression was performed to investigate the association between age and fraction contribution. Cancer types with a significant association (adj. p-value < 0.05) were further investigated using multiple linear regression. Only significant associations from multiple linear regression analysis were plotted. Circle size corresponds to the significant level, red and blue represent positive and negative associations, respectively. **c** Association between age and mutational burden in endometrial cancer after excluding tumours with MSI-H and *POLE/POLD1* mutations. Adjusted R-squared and p-value from multiple linear regression analysis are displayed.



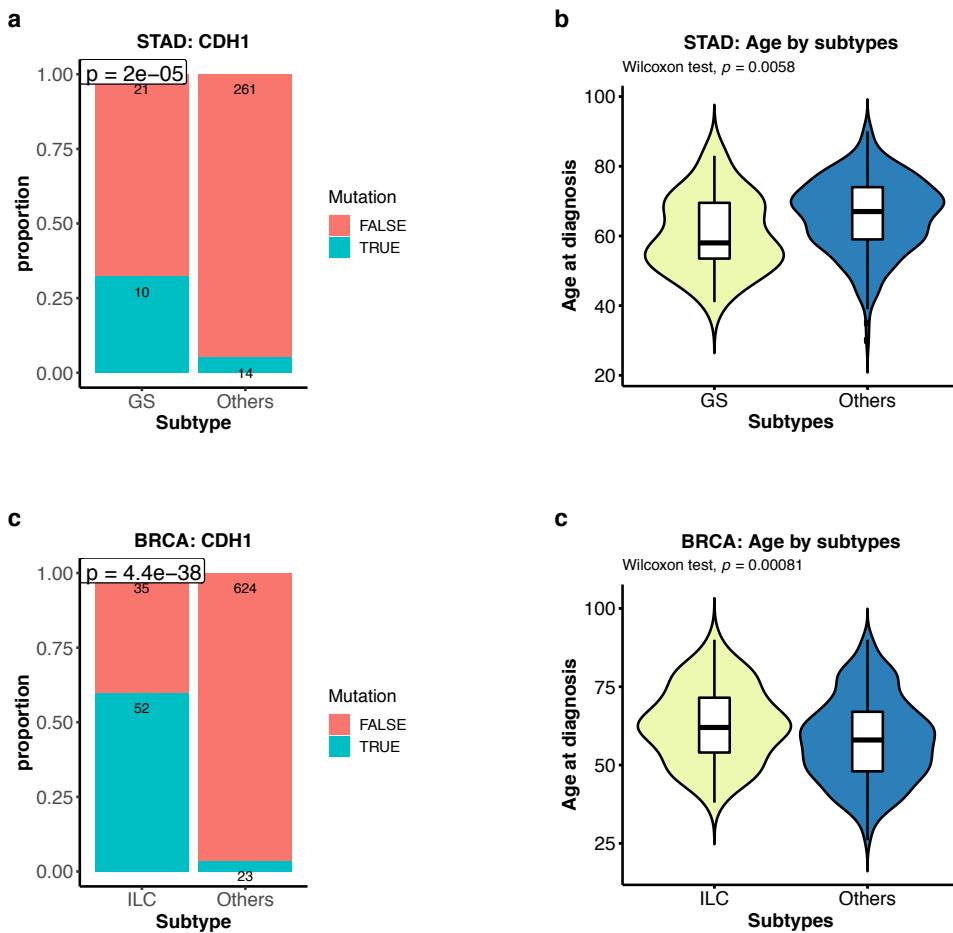
### **Supplementary Fig. 8**

Association between age and (a) MSI-H in COAD, READ, and STAD, and (b) *POLE/POLD1* mutations in cancer types containing the mutations in these genes in more than 5% of the samples. The statistical significance (p-value showed) was calculated from the multiple logistic regression adjusting for clinical variables. The middle bar of the boxplot is the median. The box represents interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of  $1.5 \times \text{IQR}$ .



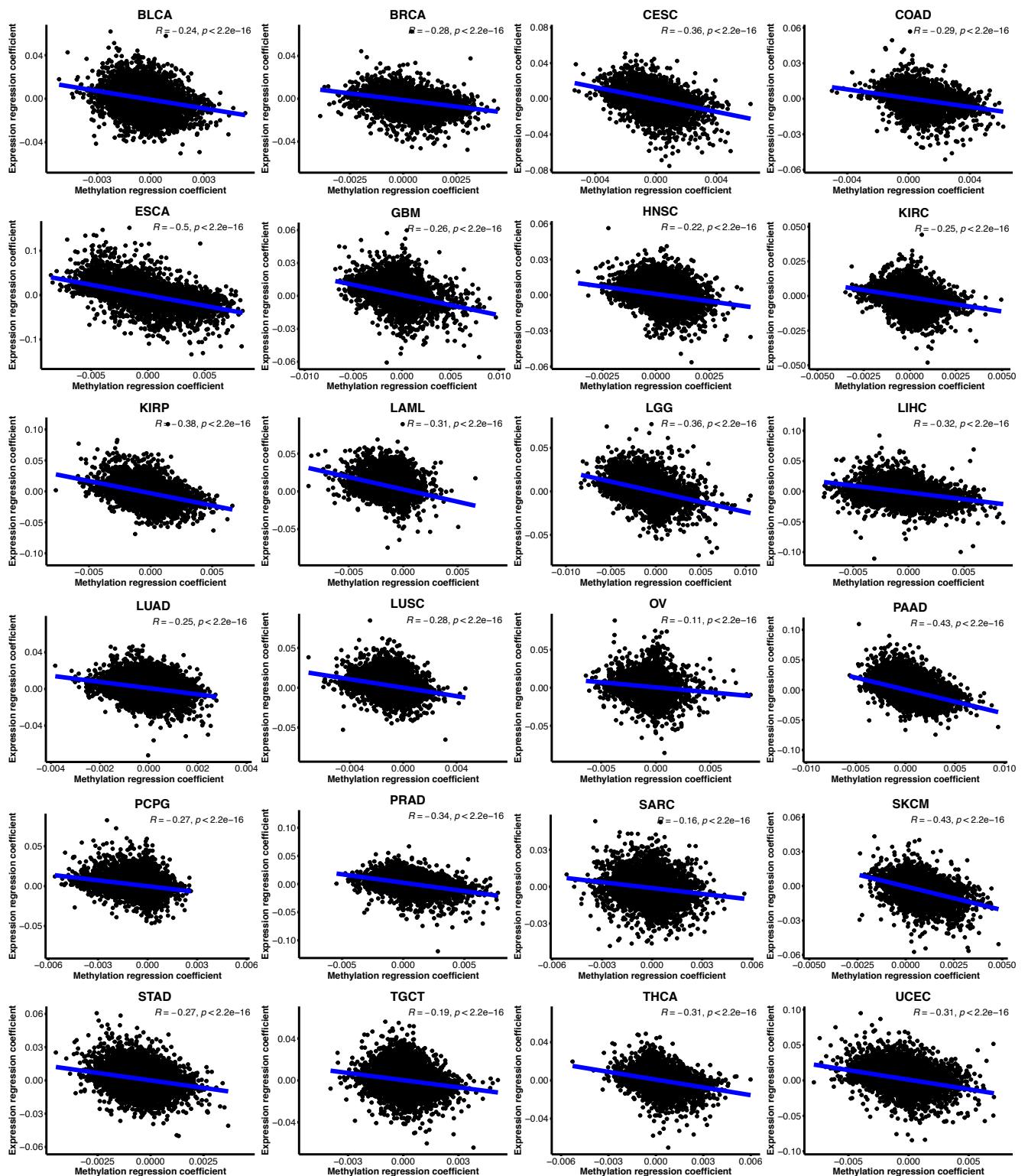
## Supplementary Fig. 9

Heatmap showing age-associated mutations in 10 cancer types. Samples are sorted by age. Colours represent types of mutation. The right annotation legend indicates the direction of change, increase or decrease mutations with age. The mutational burden of samples is presented in dots above the heatmap in a log10 scale.



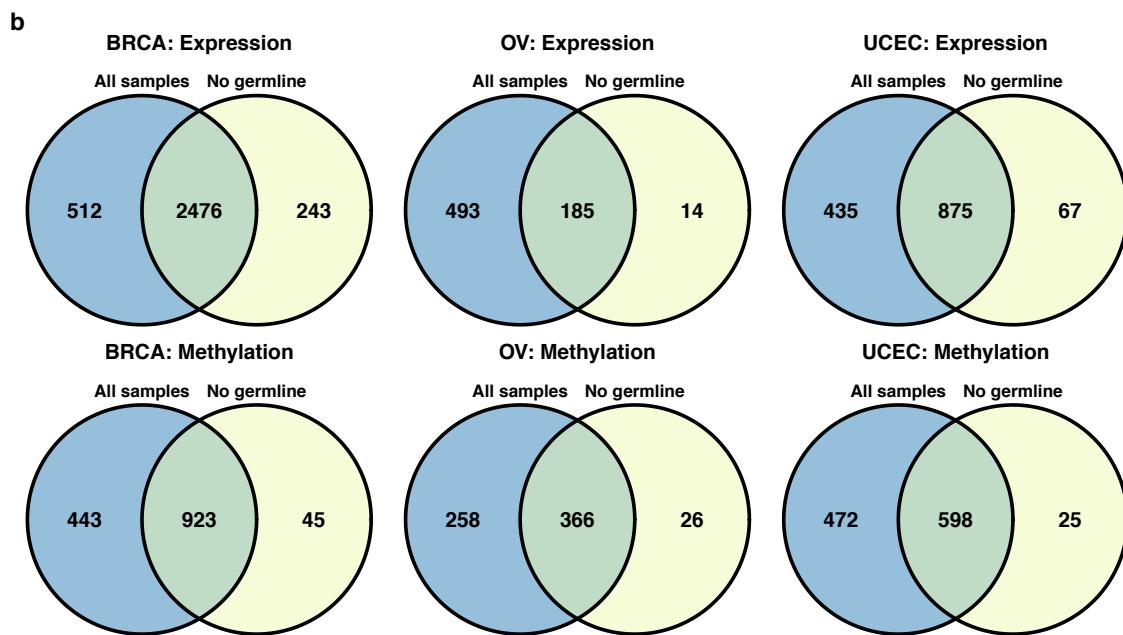
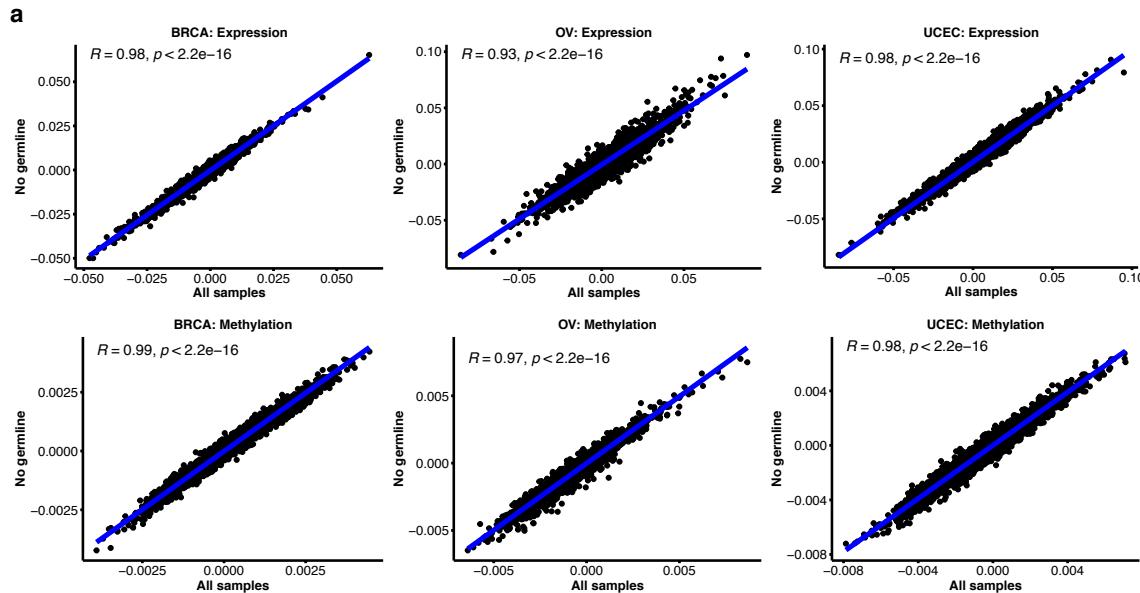
**Supplementary Fig. 10**

Mutations in *CDH1* by cancer subtypes. **a** Association between *CDH1* mutations and stomach cancer subtype comparing between genetically stable (GS) and other subtypes. The statistical significance (p-value showed) was calculated from two-sided Fisher's exact test. **b** Age distribution of genetically stable stomach cancer subtype and other stomach cancer subtypes. The statistical significance (p-value showed) was calculated from two-sided Wilcoxon rank sum test. The middle bar of the boxplot is the median. The box represents interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of  $1.5 \times$  IQR. **c** Association between *CDH1* mutations and breast cancer subtype comparing between invasive lobular carcinoma (ILC) and other subtypes. The statistical significance (p-value showed) was calculated from two-sided Fisher's exact test. **d** Age distribution of genetically stable stomach cancer subtype and other stomach cancer subtypes. The statistical significance (p-value showed) was calculated from Wilcoxon rank sum test. The middle bar of the boxplot is the median. The box represents interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of  $1.5 \times$  IQR.



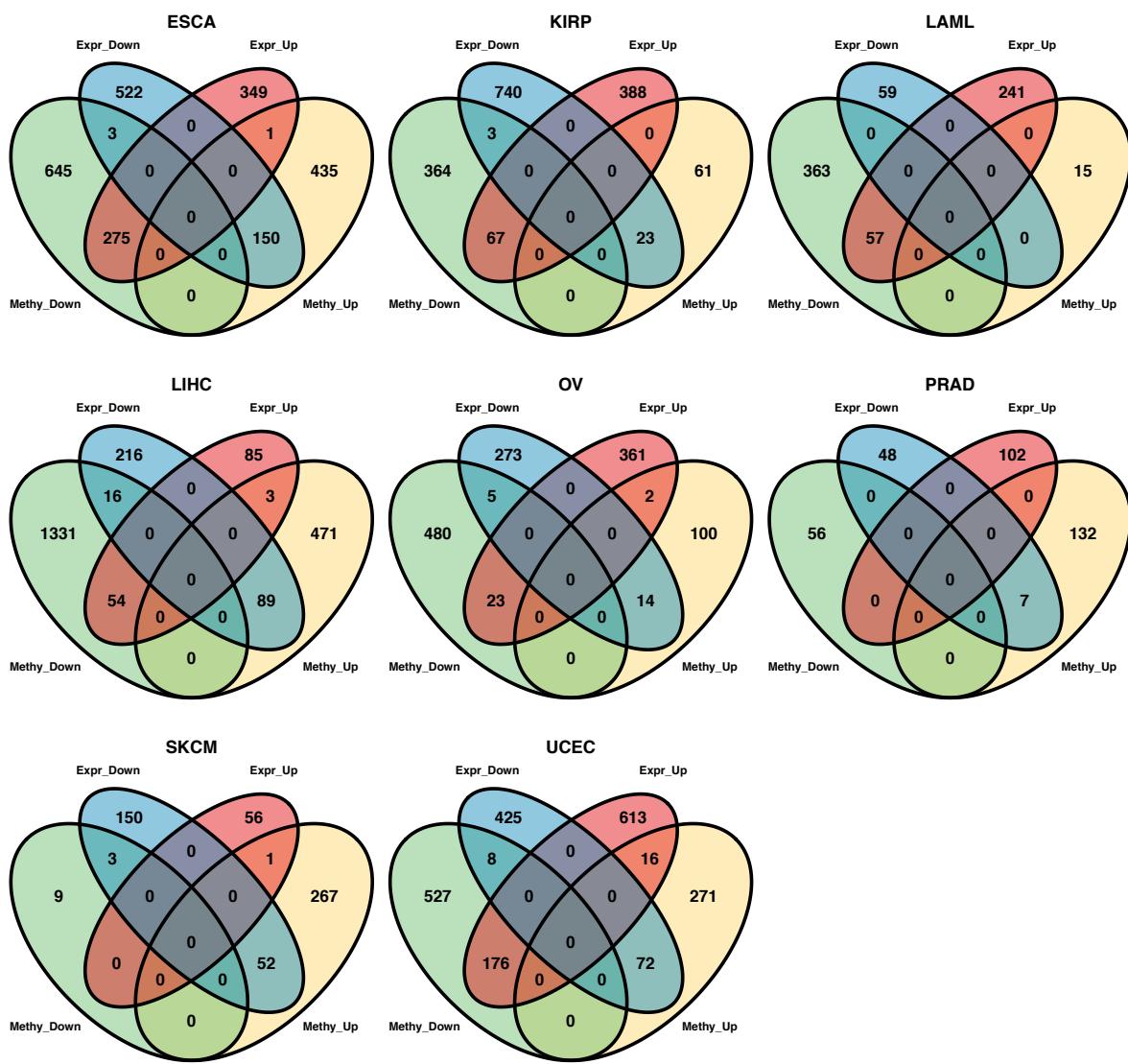
### **Supplementary Fig. 11**

Pearson correlation between linear regression coefficient of age on DNA methylation level and linear regression coefficient of age on gene expression. The regression coefficients were obtained from the multiple linear regression analysis to investigate the association between age and DNA methylation or gene expression. Pearson correlation coefficient and p-values are shown.



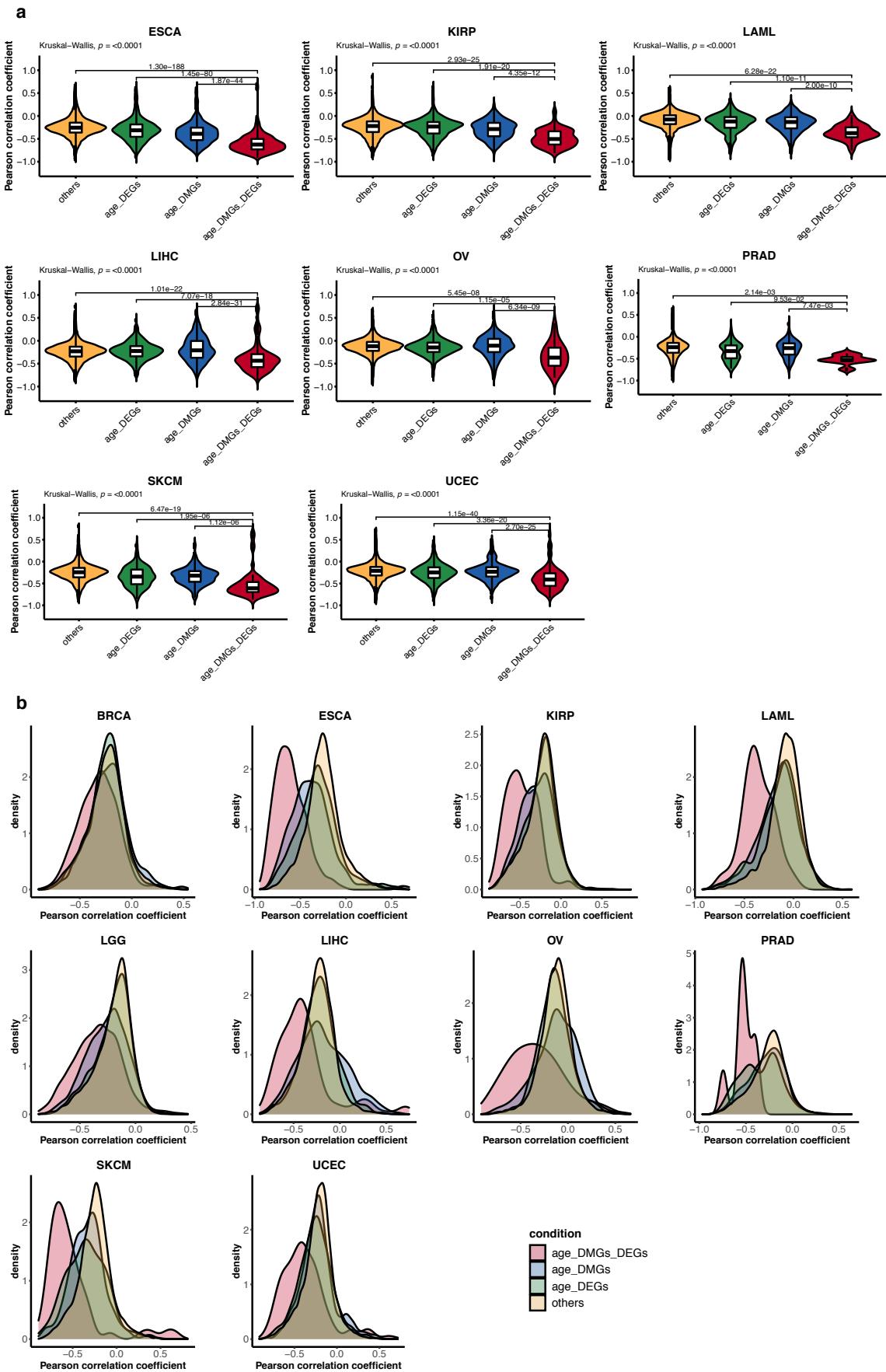
**Supplementary Fig. 12**

Impact of germline predisposition mutations on age-associated gene expression and DNA methylation in breast, ovarian and endometrial cancers. **a** Pearson correlation between linear regression coefficient of age on gene expression or methylation from all samples and only samples without germline predisposition mutations. **b** Overlap between age-DEGs or age-DMGs identified from all samples and only samples without germline predisposition mutations.



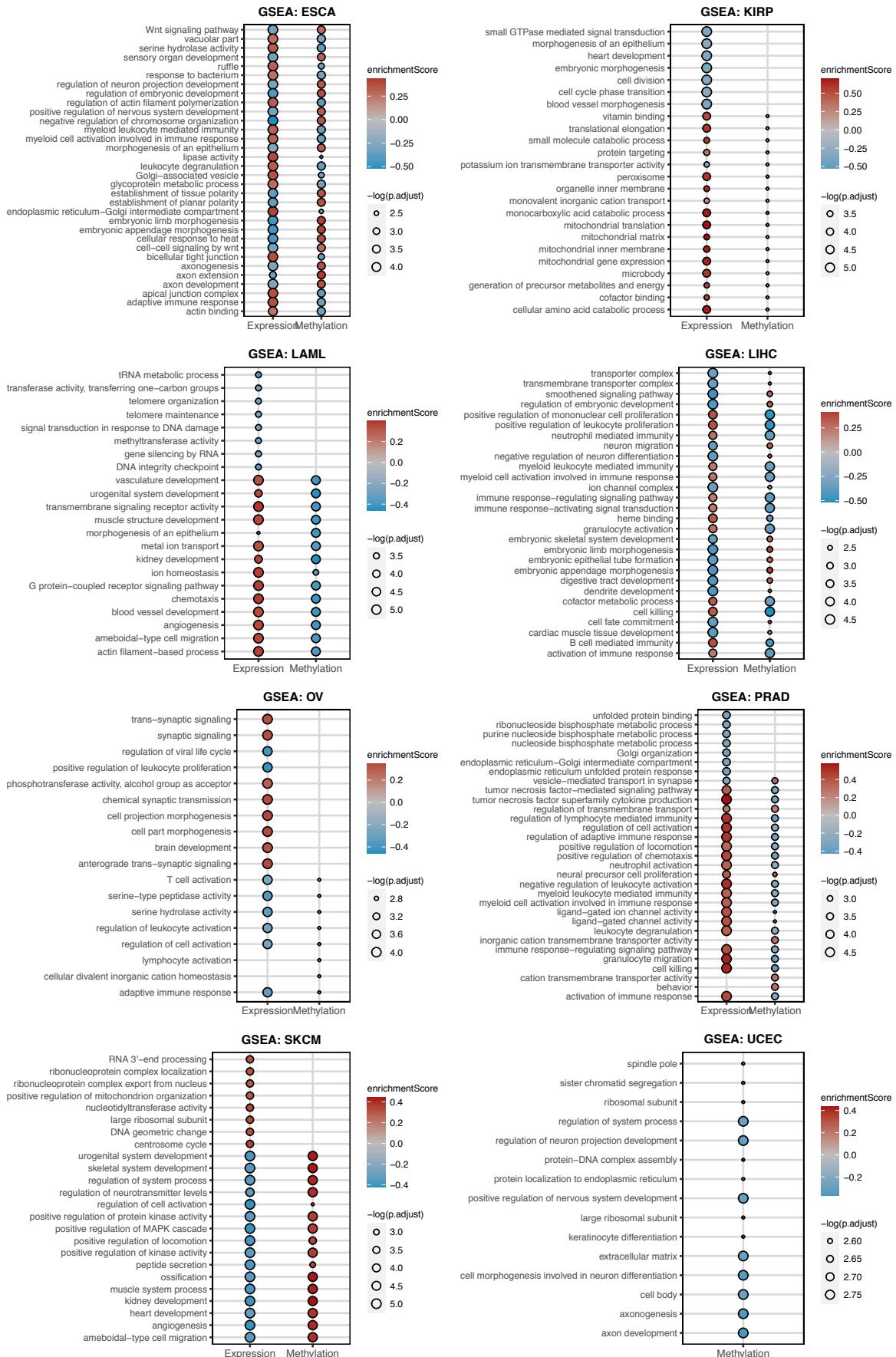
**Supplementary Fig. 13**

Venn diagrams of the overlap between age-DEGs and age-DMGs in 8 cancer types. Age-DEGs were separated into genes up-regulated with age (Expr\_Up) and genes down-regulated with age (Expr\_Down). Age-DMGs were classified into genes with increased methylation with age (Methy\_Up) and genes with decreased methylation with age (Methy\_Down).



### **Supplementary Fig. 14**

Pearson correlation coefficient between methylation and gene expression. **a** Violin plots showing the distribution of the Pearson correlation coefficient between methylation and gene expression in 8 cancer types. Genes were grouped into (1) overlapping genes between age-DMGs and age-DEGs (age-DMGs-DEGs), (2) age-DMGs only genes, (3) age-DEGs only genes, and (4) other genes. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were done using two-sided Dunn's test. P-values from Dunn's test between age-DMGs-DEGs and the other groups are shown. The middle bar of the boxplot is the median. The box represents interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of  $1.5 \times \text{IQR}$ . **b** Density plots showing the distribution of the Pearson correlation coefficient between methylation and gene expression across 10 cancer types.



### **Supplementary Fig. 15**

The enriched age-related gene ontology (GO) terms identified by GSEA in 8 cancer types. The dot size corresponds to a significant level. A GO term was considered significantly enriched if adj. p-value < 0.05 for gene expression and adj. p-value < 0.1 for methylation. Colours represent enrichment scores, red denotes positive score (enriched in older patients), while blue signifies negative score (enriched in younger patients). No enriched term was identified from the gene expression data of endometrial cancer (UCEC).