

## Figure S3

**Blood background substitution patterns exhibit previously identified signatures distinct from those in cancers**.

**a**, We focused on the amplicons in coding regions, and integrated Pan cancer somatic mutation data from exome sequencing in the TCGA to analyze patterns of base substitutions at genomic positions in the target regions which were mutated in tumor genomes and also changed in the background generated for healthy peripheral blood samples. Substitution frequency and substitution patterns were both significantly different between blood and tumors, both at highly mutated sites (mutation count > 10; Chi square test; FDR adjusted p-value <0.05) and across all such sites (Mantel test; p-value < 1e-5), with substitution patterns in tumor genomes being more skewed. It is possible that selection during cancer evolution contribute to the observed patterns. **b**, Integrating trinucleotide contexts of the substitutions, we determined the contributions of different mutation signatures previously identified in the blood gDNA background. Out of 30 previously identified signatures, our data showed overrepresentation of only 7 of them (Signatures 3, 4, 8,12, 20, 22 and 30) across different samples. Out of seven signatures, Signature 12, 3 and 4 had maximum contributions. Signature 3 and 4 are known to be associated with failure of DNA double-stranded break repair by homologous repair mechanism and tobacco mutagens respectively, whereas the aetiology of Signature 12 remains unknown. **c**, For the in the blood gDNA background, there was no systematic difference in mutation signatures between amplicons when grouped by their genomic context, and they also showed similar pattern of enrichment of few signatures as compared to others, with signature 12, 3 and 4 having maximum contributions. Signature 12 and 4 exhibits transcriptional strand bias for T>C and C>A substitutions respectively, whereas signature 3 is associated with increased numbers of large Indels.