Fig. S1. The relationship between sample mutation number and mutation signature fitting accuracy. The x-axes and y-axes show the mutation number in samples and mutational signature fitting accuracy respectively.
Fig. S2. Correlations between the number of cancer gene mutations and total mutations caused by mutational processes. The y-axes and x-axes show the numbers of cancer gene mutations and total mutations attributed to each mutational process respectively. Each panel corresponds to a mutational process in a cancer type or subtype. Each dot represents a sample. The lines represent the linear relationship calculated by robust linear regression and 95% confidence intervals for the slopes are shown in lighter gray shading.
Fig. S3. The cancer gene mutation risk of mutational processes in each cancer type. Detailed heatmap of significant cancer gene mutation risk of mutational processes in each cancer type (see Methods). The filled color of the heatmap represents the degree of risk. The risk value on the heat map is multiplied by 100 for the sake of showing it.
Fig. S4. The relationship between cancer gene mutation risk induced by SBS5 and genomic instability. (A) The relationship between cancer gene mutation risk induced by SBS5 and mutation burden. (B) The relationship between cancer gene mutation risk induced by SBS5 and copy number burden. (C) Mutation load differences among GIACs subtypes. (D) Copy number load differences among GIACs subtypes.
Fig. S5. The recurrent mutation is shaped by some mutational processes. Barplots and word clouds illustrate the recurrent mutation landscapes for SBS3, SBS18, SBS17, SBS12, SBS16, SBS22. Barplot depicts a recurrence pattern, where mutations were binned by their reoccurrence frequency and the height of the bar represented the fraction of mutations in each cancer type or subtype. The numbers of mutations in and out of cancer genes are filled by red and gray respectively and also show in parentheses separated by commas in each cancer type or subtype. Word clouds show high-frequency mutations occurring at least 6 times in a cancer type or subtype, of which word size is proportional to the number of mutations and word colored by cancer type or subtype. Note that some mutational processes are not found to cause high-frequency mutations.
Fig. S6. The sample size affected by each mutation process in each cancer type. The red numbers indicate that the mutational process causes high-frequency recurrent mutations in a cancer type.
Fig. S7. The relation between mutational processes and high-frequency mutations. The network illustrates the associations between mutational processes and high-frequency mutations. Red circles and blue rectangles correspond to mutations and mutational processes, respectively. The size of the node is proportional to the degree in the network. The thickness of the edge connecting two nodes is proportional to the number of associations in all samples.
Fig. S8. Specific mutations induced by a certain mutational process. Pie charts show the mutational process composition contributing to a mutation in a cancer type and the size is proportional to the mutation number. Significant dots (adjusted p < 0.05) are filled by the mutational process. The y coordinate of each pie center reflects mutation frequency in the corresponding cancer type; the x coordinate was determined by the normalized entropy (see Methods).
Fig. S9. Overview of identifying pathways affected by the mutational process. For a mutational process, firstly, we prioritize genes in a cancer type. Then, a ranked hypergeometric test is used to find enriched pathways. Finally, we get pathways by integrating the evidence from all cancer types. This schematic map is referenced by Paczkowska et al (Nature communication. 2020).
Fig. S10. Summary of pathways affected by mutational processes. (A) Barplot indicates the number of pathways (y-axis) affected by a certain number of the mutational process(es) (x-axis). (B) The pie chart shows the detailed composition of related mutational processes of the pathways that are specifically affected by a single mutational process.
Fig. S11. Pathways affected by clock-like mutational processes. Enrichment map of some pathways affected by A SBS1, B SBS5, and C SBS40. Nodes in the network represent pathways and are filled by cancer-type evidence. The node size indicated the number of genes in a pathway. Similar pathways with many common genes were connected.
Fig. S12. Pathways affected by APOBEC mutational process. (A) Enrichment map of some pathways affected by APOBEC mutational process solely. In this map, nodes in the network represent pathways and are colored by cancer-type evidence. The node size indicates the number of genes in a pathway. Similar pathways with many common genes were connected. (B) A scatter plot shows correlations between the log exposure value of APOBEC mutational signatures (SBS2 and SBS13, axes) and HRD signature (SBS3, y-axes). Each dot represents a sample colored according to the cancer type and the line shows best estimates for the slope estimated by mixed effect model in samples that APOBEC mutational signatures and HRD signature co-occurred. The 0 value of exposure is not processed by log transformation.
Fig. S13. HR-related genes affected by various mutational processes. The size of the pie chart represents the number of mutations caused by the mutational process in all cancer types, the color represents the different cancer types, and the area is proportional to the number of mutations in the corresponding cancer subtype.
Fig. S14. Differences between HR-Deficient and HR-Proficient groups in BRCA, CESC, and HNSC. (A) Differences in APOBEC exposure between HRD and HRP groups. (B) Differences in copy number burden between HRD and HRP groups.
Fig. S15. The contribution of the APOBEC mutation process to platinum-based therapy. Kaplan–Meier curves for different HR status and APOBEC exposure levels in (A) CESC, (B) HNSC, and (C) BRCA patients.