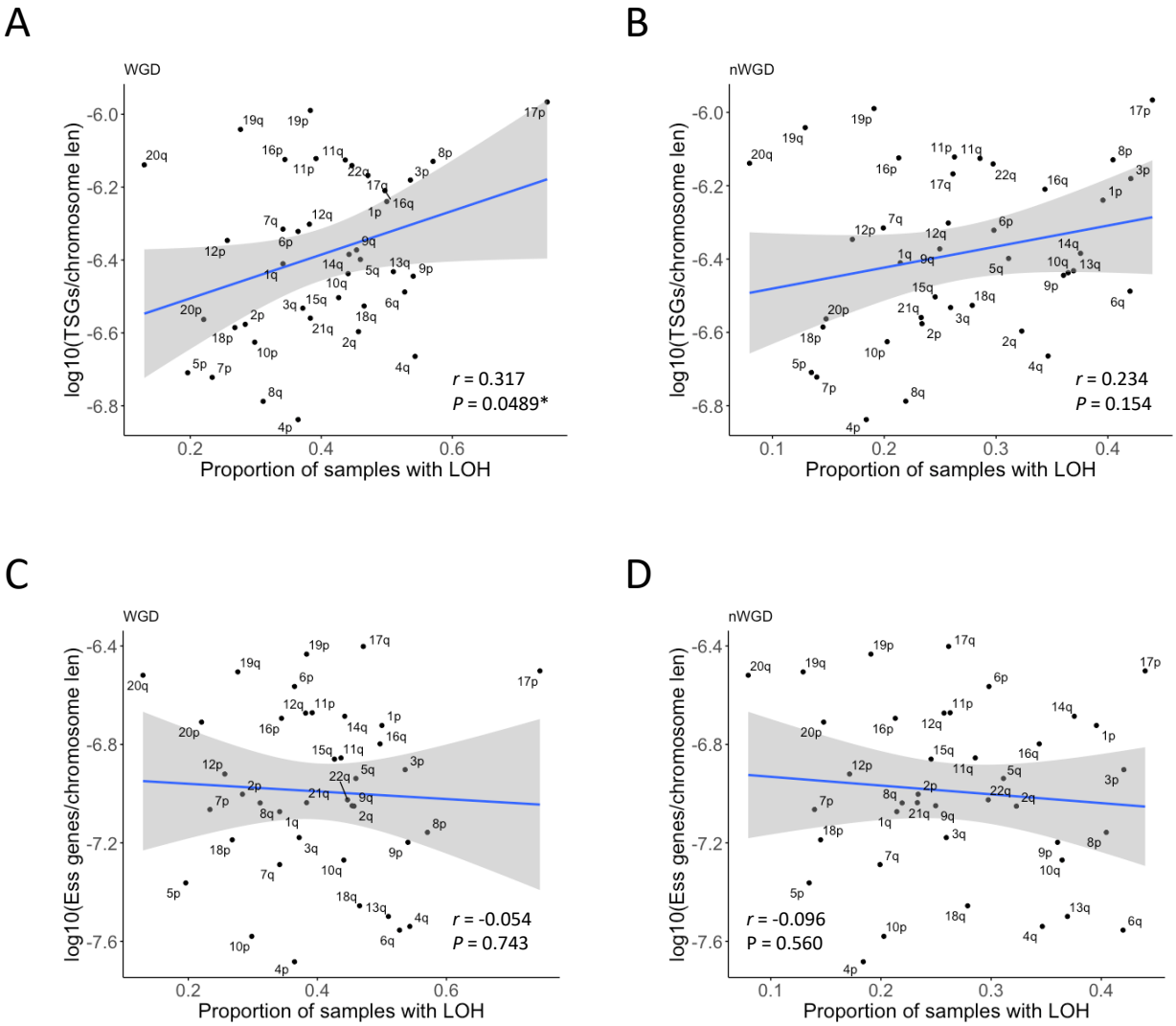
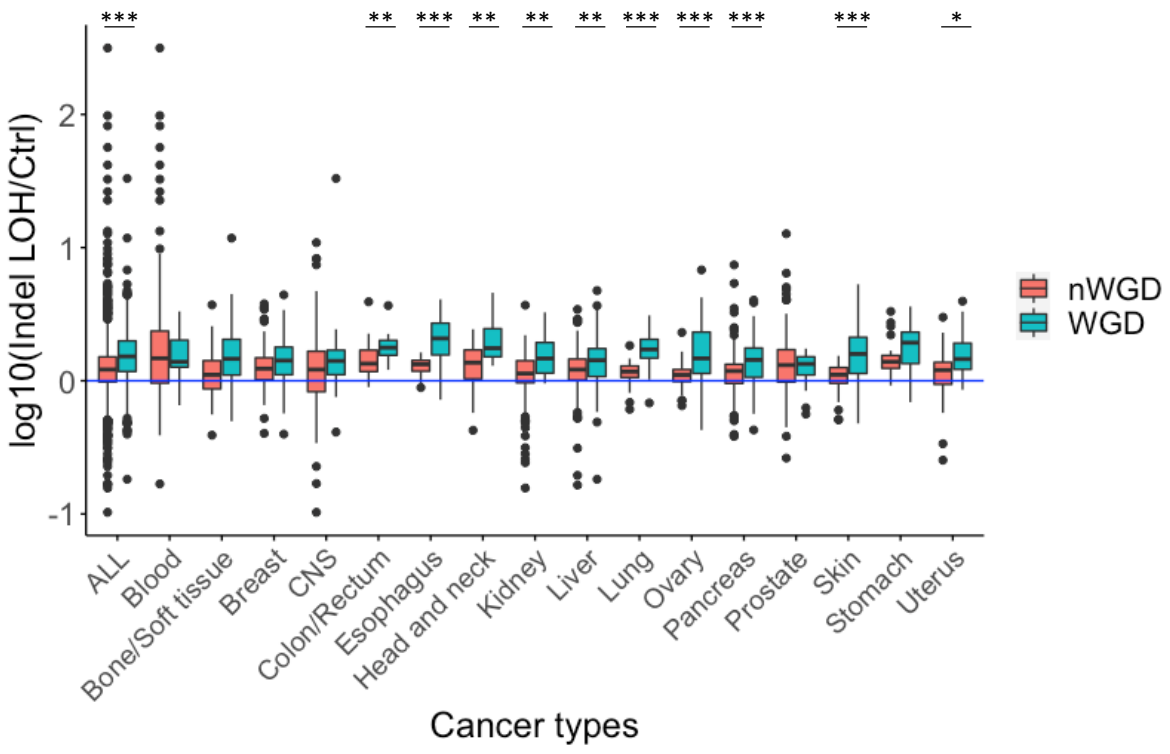


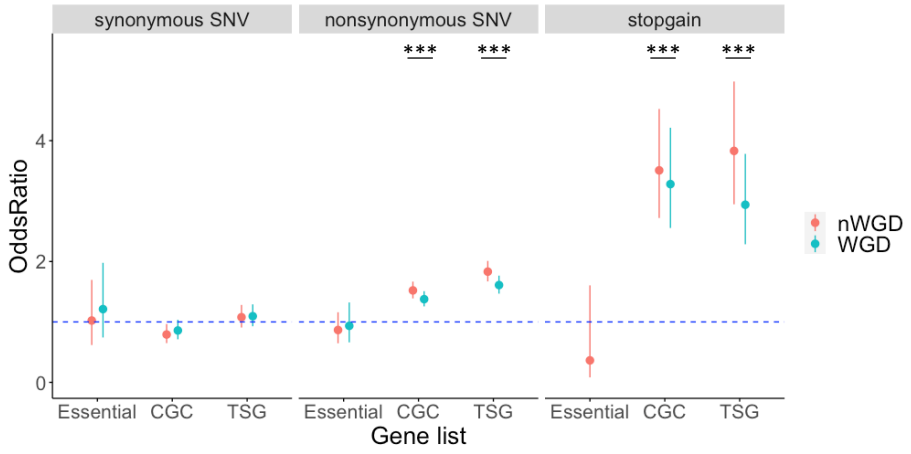
Supplementary Figure S1. The proportion of samples with LOH in each chromosome and TSG and essential gene distribution.

(A) The proportion of samples with LOH for each chromosome and log10 (number of TSGs/chromosome region length) in the WGD sample. The chromosomes were divided into long and short arms. r-correlation coefficients are indicated with P-values representing the results of testing for zero correlation. (B) The proportion of samples with LOH for each chromosome and log10 (number of TSGs/chromosome region length) in the nWGD samples. (C) The proportion of samples with LOH for each chromosome and log10 (number of essential genes/chromosome region length) in the WGD samples. (D) The proportion of samples with LOH for each chromosome and log10 (number of essential genes/chromosome region length) in the nWGD samples.



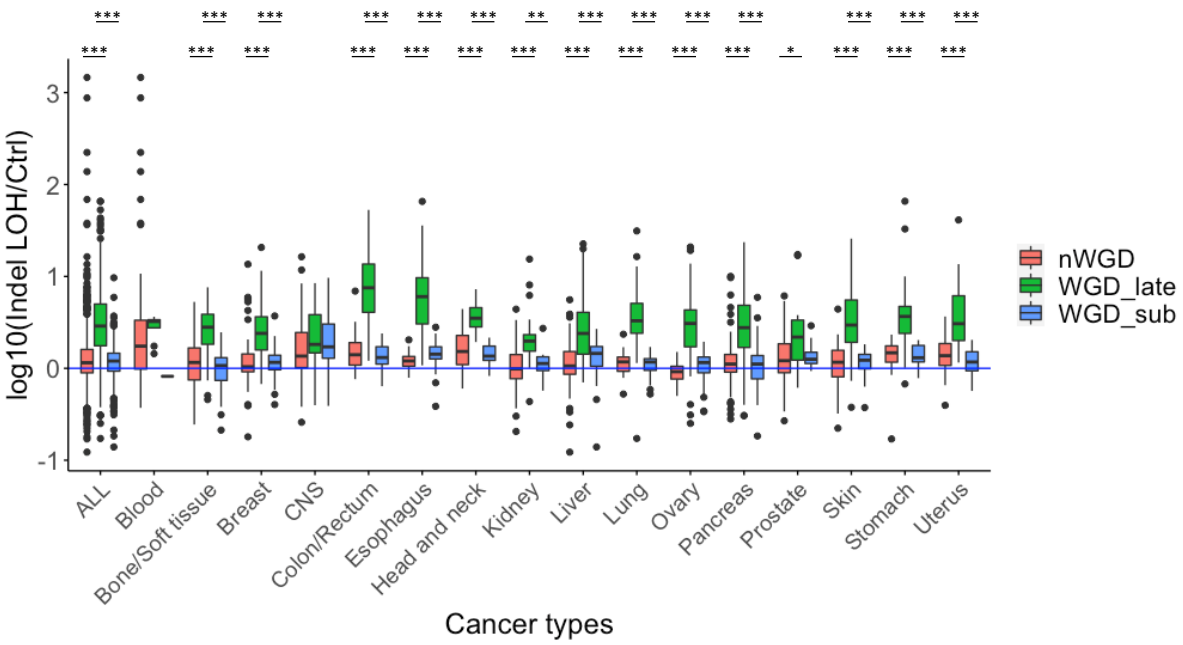
Supplementary Figure S2. The ratio of early clonal indel density within the LOH region to the density outside the LOH region in WGD samples and the ratio of clonal indel density within the LOH region to the density outside the LOH region in nWGD samples.

The vertical axis represents log10(indel density within the LOH region/indel density outside the LOH region).



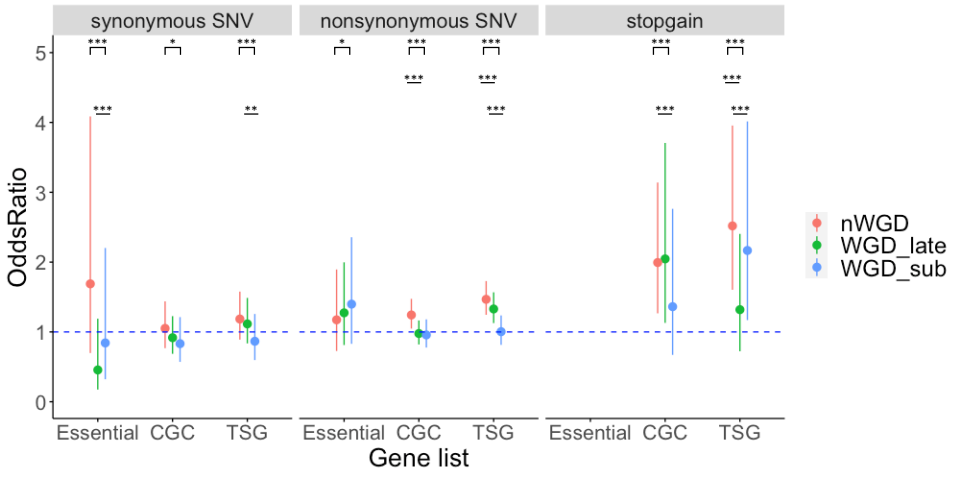
Supplementary Figure S3. Early mutation accumulation of various types in genes within the LOH region.

The blue dotted line indicates an odds ratio = 1. Odds ratios above 1 indicate that the mutation is subjected to selective pressure. Red and blue dots represent odds ratios and whiskers represent 95% confidence intervals.



Supplementary Figure S4. The ratio of late clonal and subclonal indel density within the LOH region to the density outside the LOH region in WGD samples and the ratio of subclonal indel density within the LOH region to the density outside the LOH region in nWGD samples.

The vertical axis represents log10(indel density within the LOH region/indel density outside the LOH region).

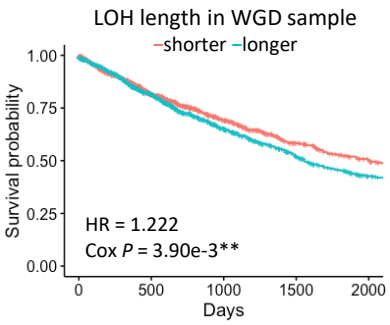


Supplementary Figure S5. Late mutation accumulation of various types in genes within the LOH region.

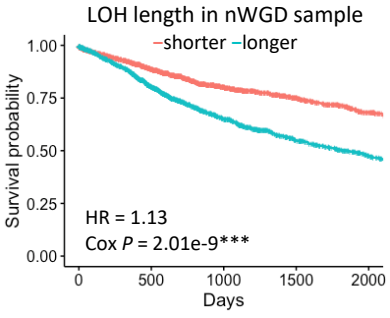
The blue dotted line indicates an odds ratio = 1. Odds ratios above 1 indicate that the mutation is subjected to selective pressure. Red and blue dots represent odds ratios and whiskers represent 95% confidence intervals.

Fig. S6
C Kikutake et. al

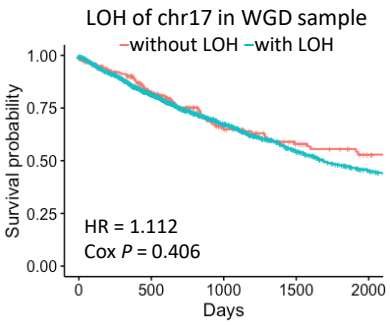
A



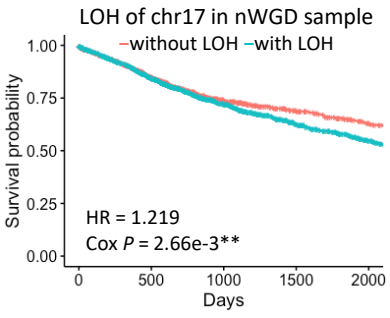
B



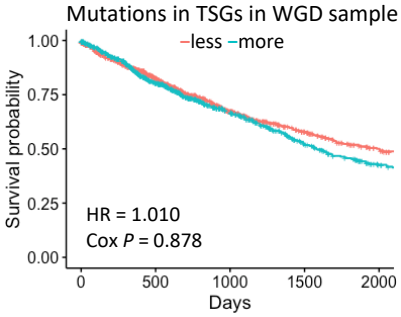
C



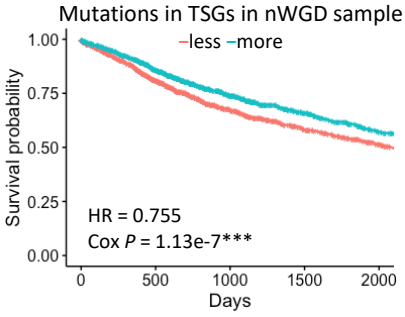
D



E



F

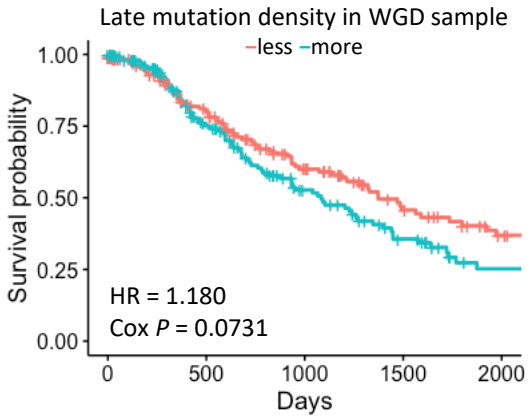


Supplementary Figure S6. Survival analysis using WGD occurrence-associated factors in the case of early mutations using TCGA data.

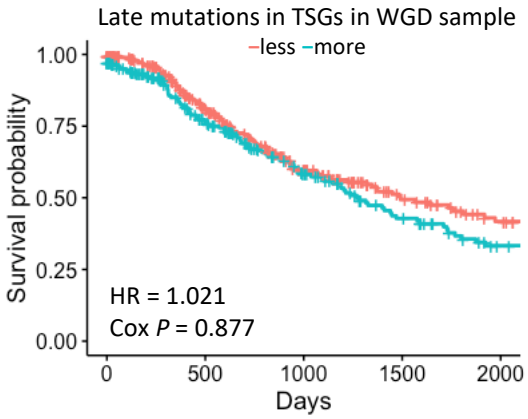
(A) Association between the LOH length and prognosis in the WGD samples using TCGA data. Samples were divided into two groups based on the median total LOH length in each sample: a longer and a shorter group (blue and red lines, respectively). The horizontal and vertical axes represent survival days (days) and probability, respectively. (B) Association between LOH length and prognosis in nWGD samples using TCGA data. The samples were divided into two groups based on the median total LOH length in each sample: a longer and a shorter group (blue and red lines, respectively). (C) Association between LOH in chr17 and prognosis in WGD samples using TCGA data. The samples were divided into two groups based on the presence or absence of LOH in chr17: groups with and without LOH (blue and red lines, respectively). (D) Association between LOH in chr17 and prognosis in nWGD samples using TCGA data. The samples were divided into two groups based on the presence or absence of LOH in chr17: groups with and without LOH (blue and red lines, respectively). (E) Association between the number of early clonal mutations in TSGs in LOH region and prognosis in WGD samples using TCGA data. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations (blue and red lines, respectively). (F) Association between the number of clonal mutations in TSGs in the LOH region and prognosis in nWGD samples using TCGA data. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations (blue and red lines, respectively).

Fig. S7
C Kikutake et. al

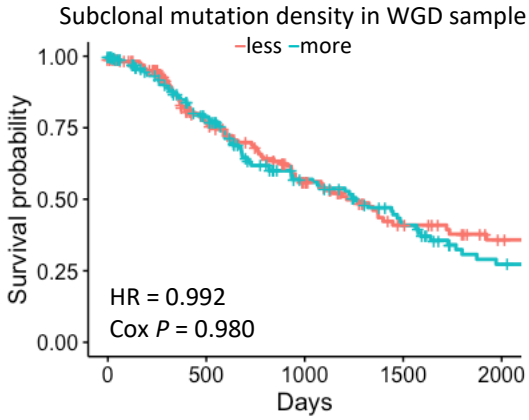
A



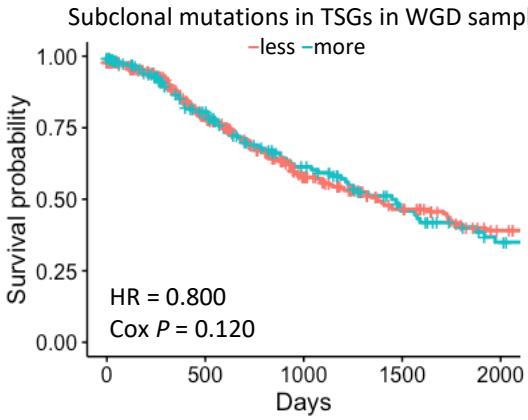
D



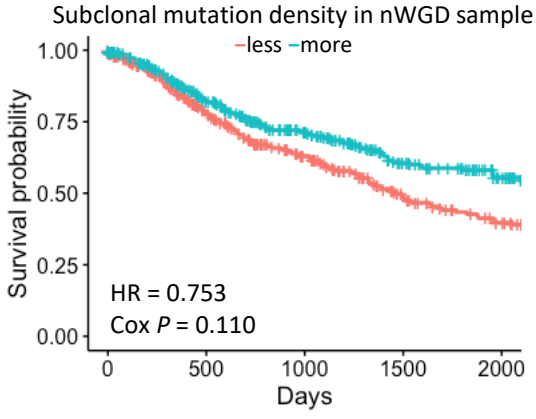
B



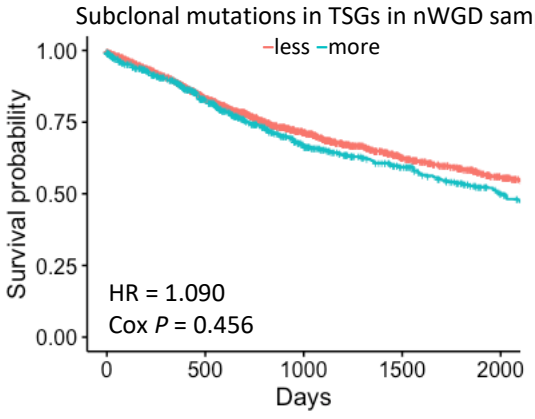
E



C



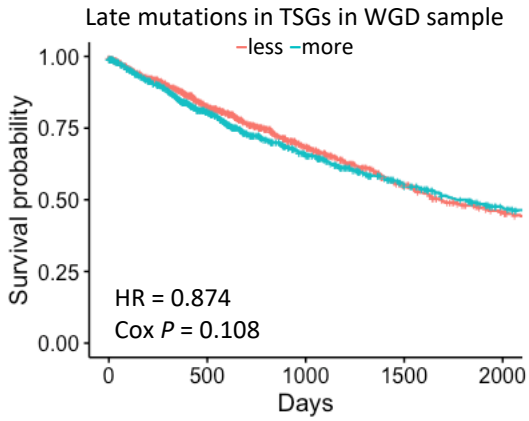
F



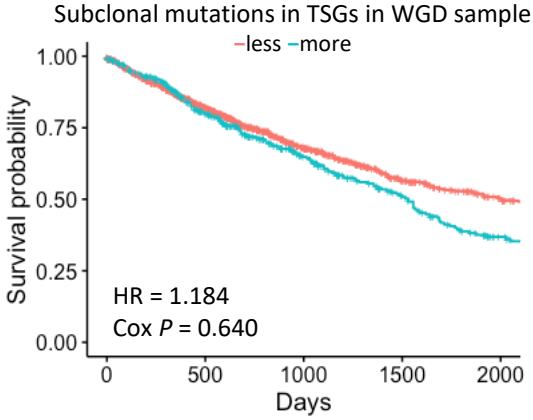
Supplementary Figure S7. Survival analysis using WGD occurrence-associated factors in the case of late mutations.

(A) Association between the ratio of late clonal mutation density within the LOH region to the density outside the LOH region and prognosis in WGD samples. The samples were divided into two groups based on the median of the ratio in each sample: a higher and a lower ratio group (blue and red lines, respectively). The horizontal and vertical axes represent survival days (days) and probability, respectively. (B) Association between the ratio of subclonal mutation density within the LOH region to the density outside the LOH region and prognosis in WGD samples. The samples were divided into two groups based on the median of the ratio in each sample: a higher and a lower ratio group (blue and red lines, respectively). (C) Association between the ratio of subclonal mutation density within the LOH region to the density outside the LOH region and prognosis in nWGD samples. The samples were divided into two groups based on the median of the ratio in each sample: a higher and a lower ratio group (blue and red lines, respectively). (D) Association between the number of late clonal mutations in TSGs in LOH region and prognosis in WGD samples. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations group (blue and red lines, respectively). (E) Association between the number of subclonal mutations in TSGs in the LOH region and prognosis in WGD samples. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations (blue and red lines, respectively). (F) Association between the number of subclonal mutations in TSGs in the LOH region and prognosis in nWGD samples. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and the less mutations (blue and red lines, respectively).

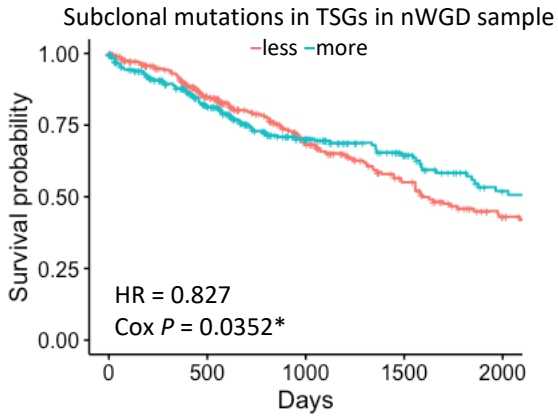
A



B



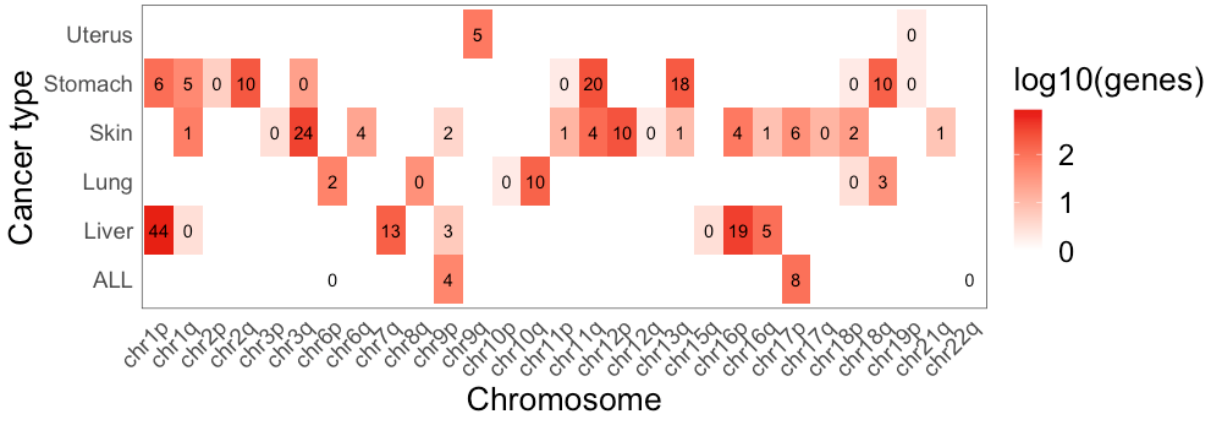
C



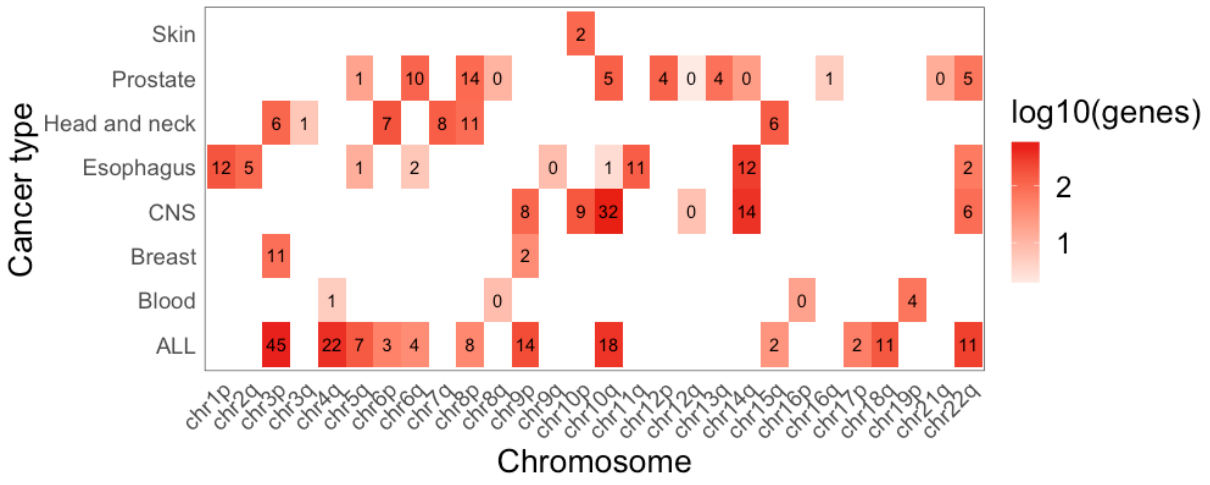
Supplementary Figure S8. Survival analysis using WGD occurrence-associated factors in the case of late mutations using TCGA data.

(A) Association between the number of late clonal mutations in TSGs in the LOH region and prognosis in WGD samples using TCGA data. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations (blue and red lines, respectively). (B) Association between the number of subclonal mutations in TSGs in the LOH region and prognosis in WGD samples using TCGA data. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations (blue and red lines, respectively). (C) Association between the number of subclonal mutations in TSGs in the LOH region and prognosis in nWGD samples using TCGA data. The samples were divided into two groups based on the median of the number of mutations in TSGs in the LOH region of each sample: groups with more and less mutations (blue and red lines, respectively).

A

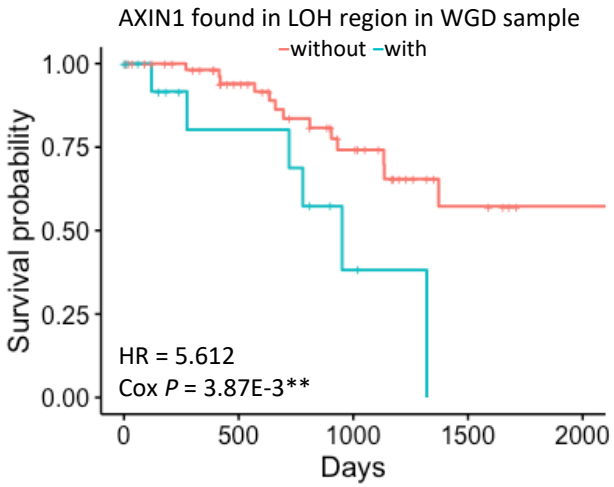


B



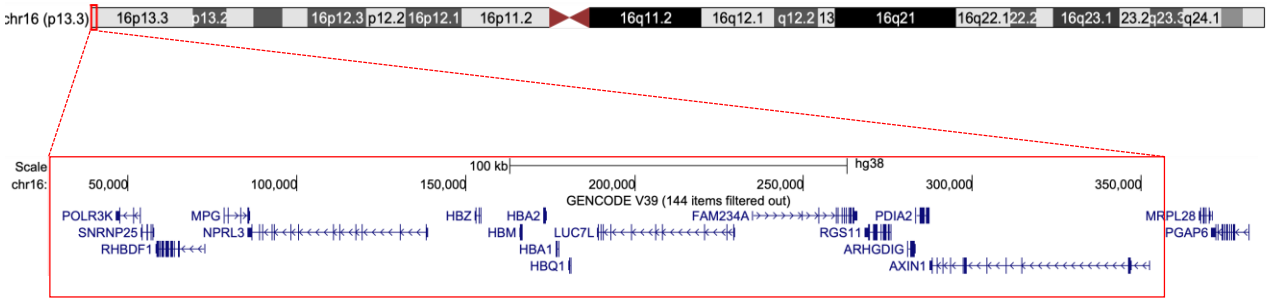
Supplementary Figure S9. Prognostic gene exploration using the PCAWG data.

(A) The number of prognosis-related genes located in LOH in WGD samples using the PCAWG data. These genes represent those that showed a worse prognosis when found in LOH region than when not found in LOH region. These genes represent those that showed a worse prognosis when found in LOH region than when not found in LOH region. The red square indicates the number of genes. Numbers in the squares represent the TSG numbers of the genes. (B) The number of prognosis-related genes located in LOH in nWGD samples using the PCAWG data.



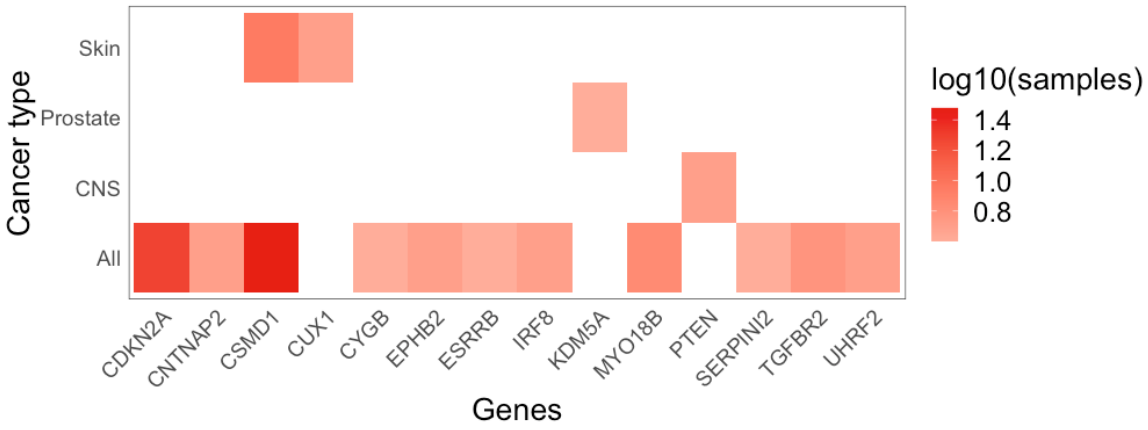
Supplementary Figure S10. Association between *AXIN1* in LOH and prognosis in WGD samples using the PCAWG data.

The samples were divided into two groups based on the presence or absence of *AXIN1* in LOH: samples with *AXIN1* in LOH region (blue line) and samples without *AXIN1* in LOH region (red line). The horizontal and vertical axes represent survival days (days) and probability, respectively.



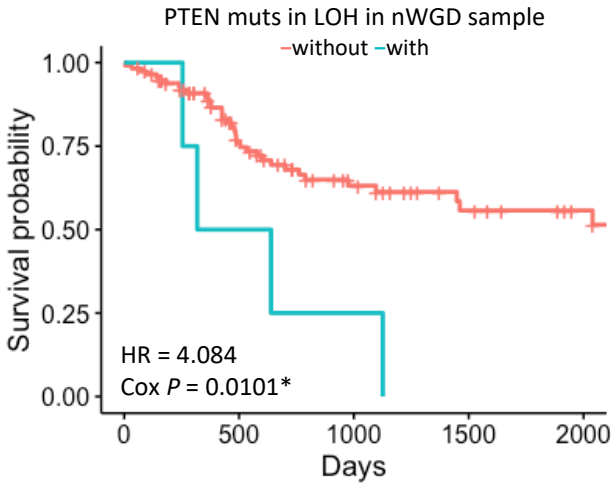
Supplementary Figure S11. Extracted prognosis-related genes in the 16p.13.3 locus.

The red square represents the 16 extracted prognosis-related genes in the 16p.13.3 locus.



Supplementary Figure S12. Number of samples with prognosis-related TSGs in LOH of the WGD samples using the PCAWG data.

The samples with the mutations in these genes found in LOH represent a worse prognosis than other samples. The red square indicates the number of samples with mutations in genes found in LOH.



Supplementary Figure S13. Association between *PTEN* mutations in LOH and prognosis in nWGD samples using the PCAWG data.

The samples were divided into two groups based on the presence or absence of mutations in *PTEN* found in LOH: samples with mutations in *PTEN* (blue line) and samples without mutations in *PTEN* (red line).