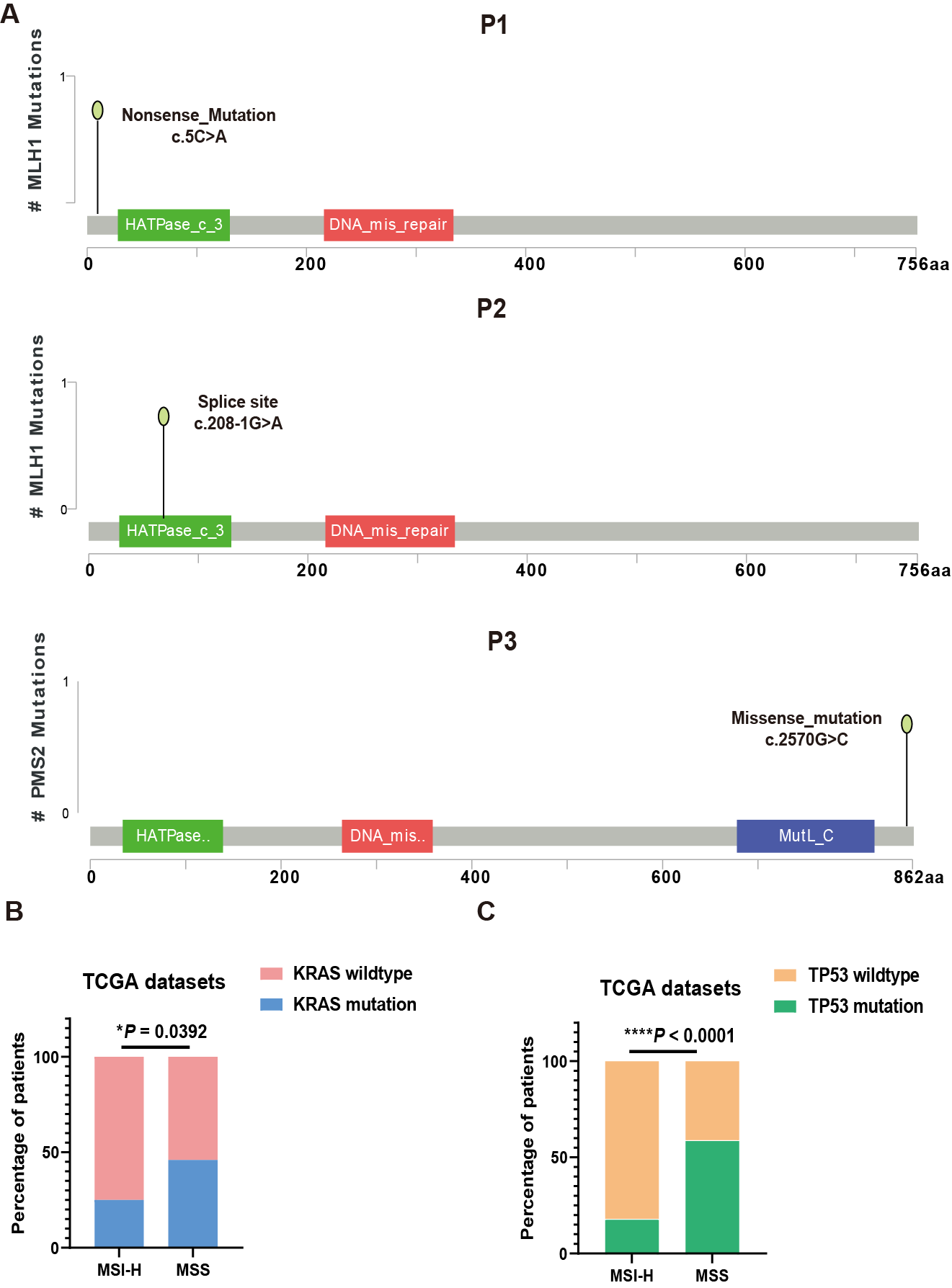
**Additional files**

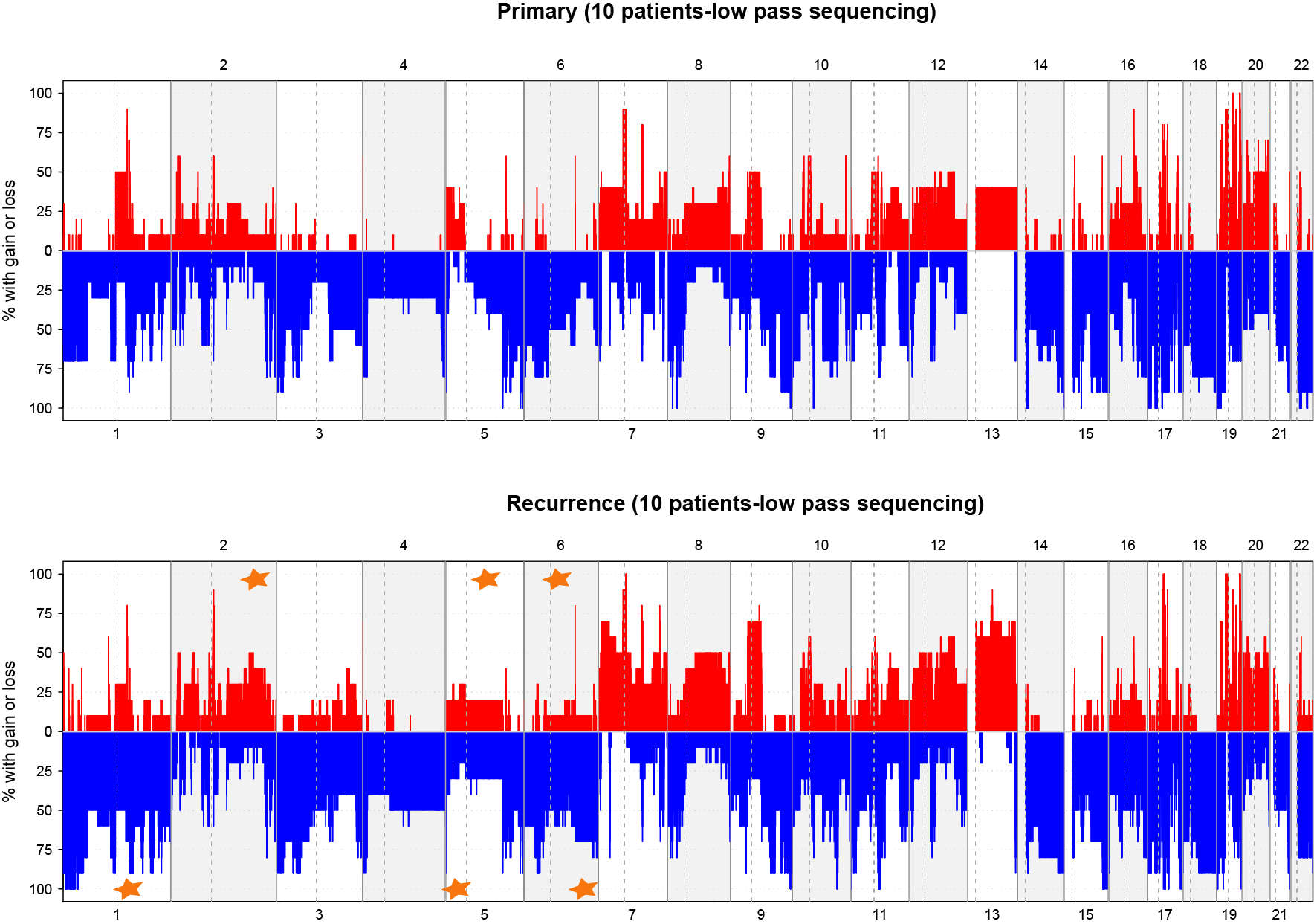


**Additional Figure 1.** The germline variations in DNA mismach repair genes, and the distribution of *KRAS* and *TP53* mutation in TCGA dataset. (A) Germline variations in mismatch repair genes for P1 to P3 (cbioportal). Distributions of (B) *KRAS* mutations and (C) *TP53* mutations in TCGA dataset. Fisher’s exact test, \*P=0.0392, \*\*\*\*P<0.0001.

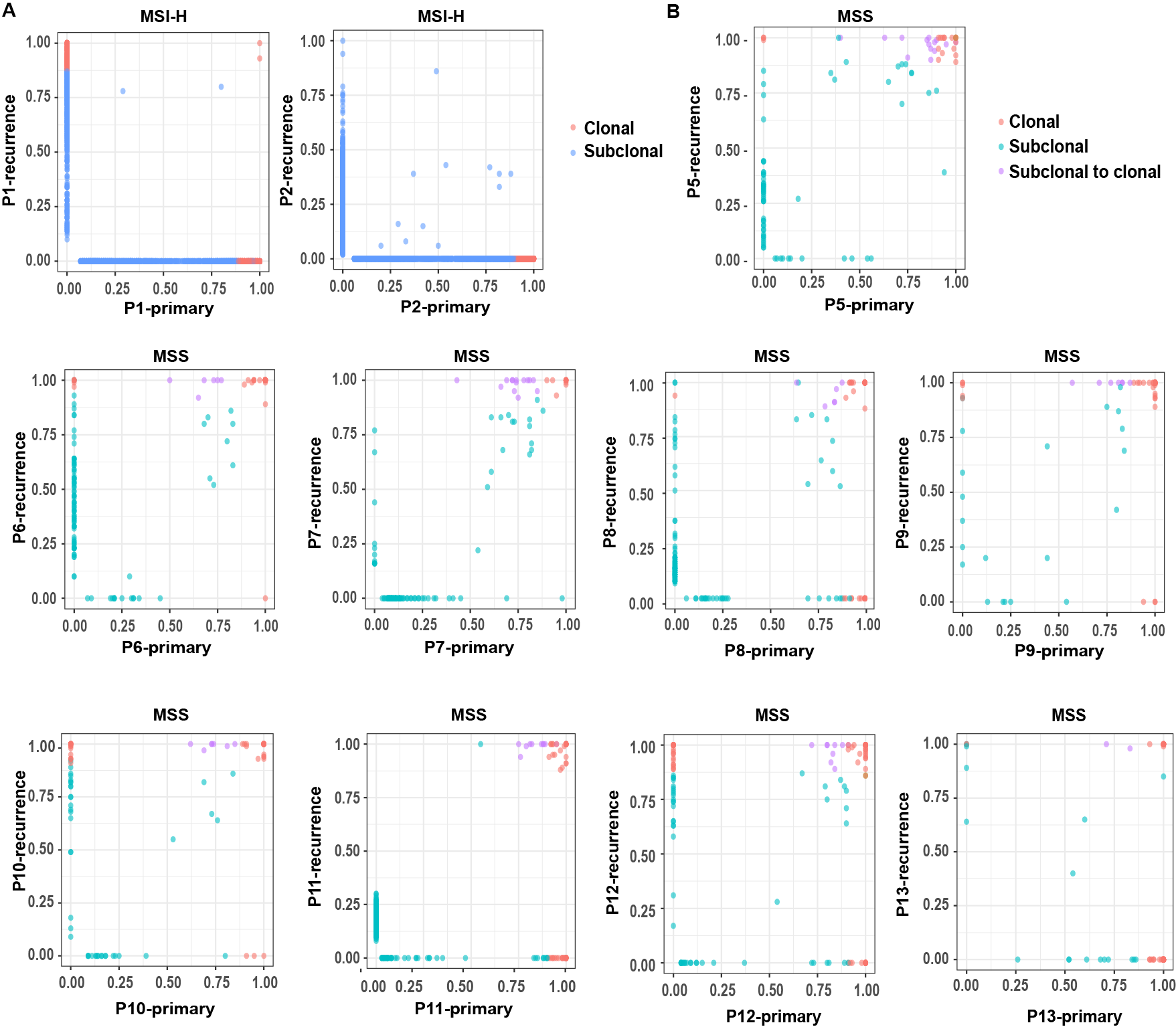
图表

描述已自动生成

**Additional Figure 2**. CNVs for MSI-H patients (P1–P3). The less frequent CNV was presented in MSI-H patients. For each patient, the top panel indicates tangent-normalized copy ratio, and the bottom panel indicates minor allele frequency. CNV=Copy number variation, MSI-H=high microsatellite instability.



**Additional Figure 3.** Validation dataset of fragment variations between primary and LR lesions. X axis indicates chromosome number and Y axis indicates percentage of copy gain or loss; the red indicates copy gain of the fragment; The blue indicates copy loss of the fragment; the star indicates the significant gain or loss in LR lesions compare with primary lesions. LR=locoregional recurrence.



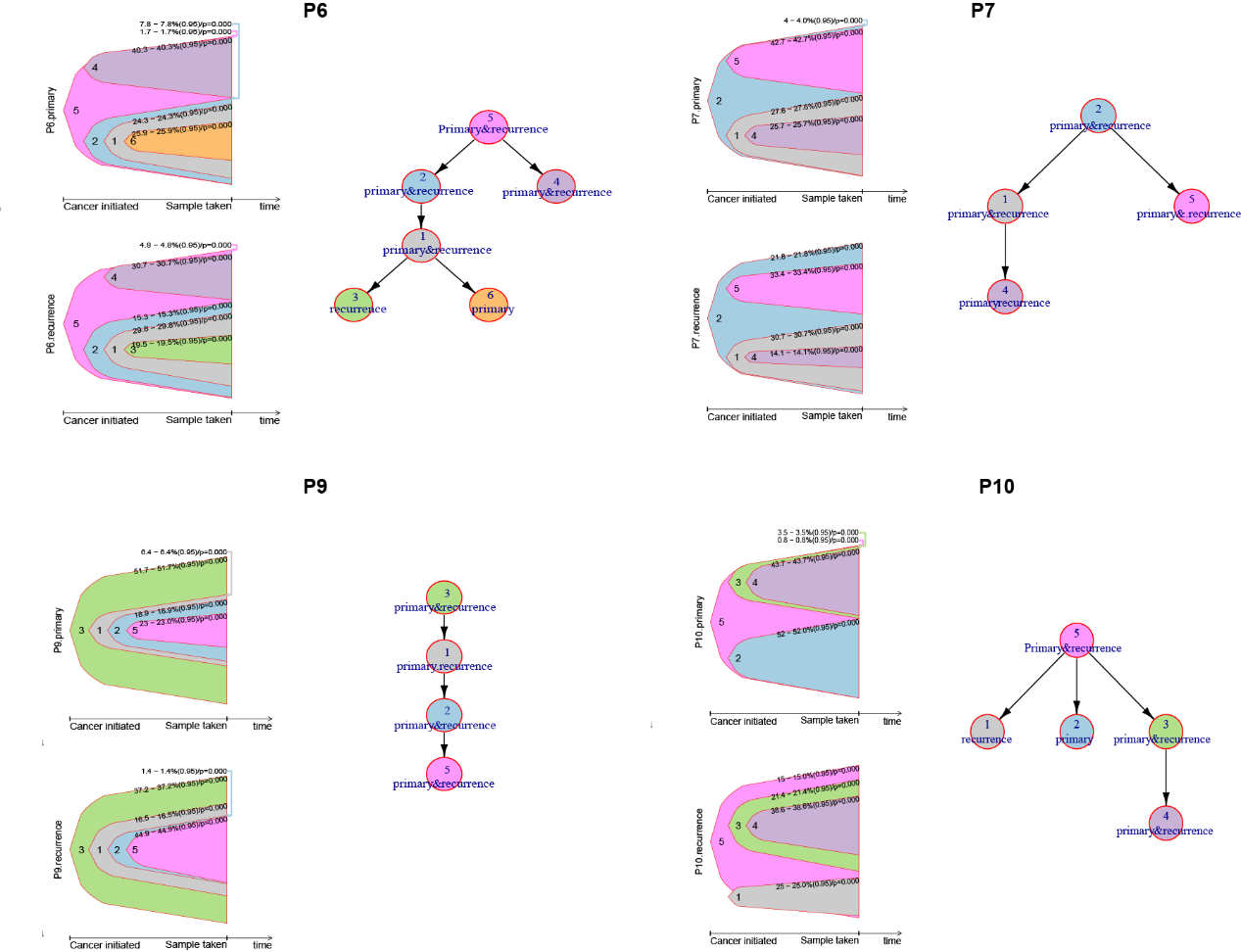
**Additional Figure 4.** CCF dynamics in (A) MSI-H and (B) MSS patients. The dots represent mutations in each patient. X axis indicates the CCF of primary lesions and Y axis indicates the CCF of LR lesions. CCF=cancer cell fraction, LR=locoregional recurrence, MSI-H=high microsatellite instability, MSS=microsatellite stable.



**Additional Figure 5.** The distribution of exclusive mutations in LR lesions and the pathway enrichment of these genes. (A) LR-specific mutation excess for each MSS patient. (B) Correlation between recurrence-specific mutation excess and time interval of DFS. (C, D) Significantly enriched pathways for mutations exclusive to LR lesions. DFS=disease free survival, LR=locoregional recurrence, MSS=microsatellite stable.



**Additional Figure 6**. Clusters of disseminating cells. The VAF dynamic of each MSS patient. Each line and dot pair represents the same disseminating cluster. MSS=microsatellite stable, VAF=variant allele frequency.



**Additional Figure 7.** Evolutionary models for representative patients. For each figure, the left panel indicates the clone evolved process in primary (top) and LR (bottom) lesion; The right panel indicates the hierarchical of each clone during the evolve process. LR=locoregional recurrence.



**Additional Figure 8.** The comparison of mutational characteristics for early- and late-recurrenced patients.(A, B) Significantly enriched pathways for patients with early and late recurrence. (C) Comparison of DFS between ID8-positive and -negative patients (ns, *P*=0.418; Wilcoxon test). (D) Comparison of DFS between ID4- or ID8-positive and double-negative patients (ns, *P*=0.094; Wilcoxon test). DFS=disease free survival, ID=indel, ns=not significant.