Supplementary Figure S1. Mutation identified by in tumor and adjacent normal tissues for each patient. X-axis: sequencing depth. Y-axis: allele fraction of mutation. Blue dots: mutations identified in tumor tissues. Red dots: mutations identified in normal tissues that most represent amplification or sequencing errors.
Supplementary Figure S2. The relationship of clonal and subclonal mutations along with normal tissue contamination. The black oval represents the clonal mutations which occupy the entire tumor, and the red and blue ovals represent two subclones which occupy only a subset of tumor cells. The sizes of the clonal and subclonal ovals represent the relative CCFs, with a theoretical value of CCF=1 for clonal mutations (the estimates are always lower given the estimate uncertainties, e.g., CCF=0.8 is used in the current study). Some normal cells are also included here to represent contamination from normal cells in the real data.
Supplementary Figure S3. The inference of subclonal architecture based on estimated CCFs in multiple blocks of primary and metastatic tumors. The colors represent clonal or subclonal mutations, with their sizes representing the estimated CCFs. The theoretical upper bound of clonal mutations should be close to 1, and the relationship of subclones can be deduced based on the relative size of the subclones. The detailed principles are described in Methods.
Supplementary Figure S4. The distribution of the 96 trinucleotide somatic mutations types. The 96 substitution classification is defined by the substitution class and sequence context immediately 3′ and 5′ to the mutated base. The x-axis represents 96 somatic mutations types. There are six classes of base substitutions, including C->A (black), C->G (red), C->T (green), T->A (blue), T->C (cyan), T->G (pink). The y-axis represents the number of each of the 96 trinucleotide somatic mutations types.
Supplementary Figure S5. The profiles of read depth ratios and allele ratios at the germline heterozygous variants for each patient. Panel A) shows the copy number alterations (CNAs) of each tumor region. The y-axis represents the log ratios of tumor read depth versus normal read depth and the x-axis represents chromosomes. HEMD (green) is a hemizygous deletion, NEUT (blue) is a copy neutral, and GAIN (red) is an amplification. Panel B) shows the allele ratios (AR) as the proportion of reads matching the reference genome. The y-axis represents the allele ratios (#reference bases/#total depth), and the x-axis represents chromosomes. HET (gray) is heterozygous, LOH (green) is loss of heterozygous, NLOH (blue) is copy neutral LOH, and ASCNA(red) is allele-specific amplification.
Supplemental Figure S6. The relationship of CCFs for all mutations across all tumor regions for each patient. Colors are the same as those in Figure 1.
Supplemental Figure S7. The clusterings of mutations based on their CCFs in different regions of patient A01
Supplementary Figure S8. The clusterings of mutations based on their CCFs in different regions of patient A02, A03, and A04. The clusterings were obtained using Pyclone by jointly modeling mutations in all tumor samples of this patient. The x-axis represents the CCFs of the mutations and the y-axis represents the somatic mutations and their corresponding genes.